Novel cell cycle genes and uses thereof

The present invention relates to DNA sequences encoding cell cycle interacting proteins as well as to methods for obtaining the same. The present invention also provides vectors comprising said DNA sequences, wherein the DNA sequences are operatively linked to regulatory elements allowing expression in prokaryotic and/or eukaryotic host cells. In addition, the present invention relates to the proteins encoded by said DNA sequences, antibodies to said proteins and methods for their production. Furthermore, the present invention relates to regulatory sequences which naturally regulate the expression of the above described DNA sequences. The present invention also relates to a method for controlling or altering growth characteristics of a plant and/or a plant cell comprising introduction and/or expression of one or more cell cycle regulatory proteins functional in a plant or parts thereof and/or one or more DNA sequences encoding such proteins. Also provided by the present invention is a process for disruption plant cell division by interfering in the expression of a substrate for cyclin-dependent protein kinase using a DNA sequence according to the invention wherein said plant cell is part of a transgenic plant. The present invention further relates to diagnostic compositions comprising the aforementioned DNA sequences, vectors, proteins and antibodies. The present invention also relates to methods for the identification of compounds being capable of activating or inhibiting the cell cycle. Furthermore, the present invention relates to transgenic plant cells, plant tissue and plants containing the above-described DNA sequences, regulatory sequences and vectors as well as to the use of the aforementioned DNA sequences, regulatory sequences, vectors, proteins, antibodies and/or compounds identified by the method of the invention in plant cell and tissue culture, plant breeding and/or agriculture.

Several documents are cited throughout the text of this specification. Each of the documents cited herein (including any manufacturer's specifications, instructions etc.)

are hereby incorporated herein by reference; however, there is no admission that any document cited is indeed prior art as to the present invention.

Cell division is fundamental for growth in humans, animals and plants. Prior to dividing in two daughter cells, the mother cell needs to replicate its DNA. The cell cycle is traditionally divided into 4 distinct phases:

G1: the gap between mitosis and the onset of DNA synthesis;

S: the phase of DNA synthesis;

G2: the gap between S and mitosis;

M: mitosis, the process of nuclear division leading up to the actual cell division.

The distinction of the 4 cell cycle phases provides a convenient way of dividing the interval between successive divisions. Although they have served a useful purpose, a recent flurry of experimental results, much of it as a consequence of cancer research, has resulted in a more intricate picture of the cell cycle's "four seasons" (Nasmyth, Science 274, 1643-1645, 1996; Nurse, Nature, 344, 503-508, 1990). The underlying mechanism controlling the cell cycle control system has only recently been studied in greater detail. In all eukaryotic systems, including plants, this control mechanism is based on two key families of proteins which regulate the essential process of cell division, namely protein kinases (cyclin dependent kinases or CDKs) and their activating associated subunits, called cyclins. The activity of these protein complexes is switched on and off at specific points of the cell cycle. Particular CDK-cyclin complexes activated at the G1/S transition trigger the start of DNA replication. Different CDK-cyclin complexes are activated at the G2/M transition and induce mitosis leading to cell division. Each of the CDK-cyclin complexes execute their regulatory role via modulating different sets of multiple target proteins. Furthermore, the large variety of developmental and environmental signals affecting cell division all converge on the regulation of CDK activity. CDKs can therefore be seen as the central engine driving cell division.

In animal systems and in yeast, knowledge about cell cycle regulations is now quite advanced. The activity of CDK-cyclin complexes is regulated at five levels: (i) transcription of the CDK and cyclin genes; (ii) association of specific CDKs with their specific cyclin partner; (iii) phosphorylation/dephosphorylation of the CDK and cyclins; (iv) interaction with other regulatory proteins such as SUC1/CKS1 homologues and cell

cycle kinase inhibitors (CKI); and (v) cell cycle phase-dependent destruction of the cyclins and CKIs.

The study of cell cycle regulation in plants has lagged behind that in animals and yeast. Some basic mechanisms of cell cycle control appear to be conserved among eukaryotes, including plants. Plants were shown to also possess CDK's, cyclins and CKIs. However plants have unique developmental features which are reflected in specific characteristics of the cell cycle control. These include for instance the absence of cell migration, the formation of organs throughout the entire lifespan from specialized regions called *meristems*, the formation of a cell wall and the capacity of non-dividing cells to re-enter the cell cycle. Another specific feature is that many plant cells, in particular those involved in storage (e.g. endosperm), are polyploid due to rounds of DNA synthesis without mitosis. This so-called endoreduplication is intimately related with cell cycle control.

Due to these fundamental differences, multiple components of the cell cycle of plants are unique compared to their yeast and animal counterparts. For example, plants contain a unique class of CDKs, such as CDC2b in *Arabidopsis*, which are both structurally and functionally different from animal and yeast CDKs.

The further elucidation of cell cycle regulation in plants and its differences and similarities with other eukaryotic systems is a major research challenge. Strictly for the case of comparison, some key elements about yeast and animal systems are described below in more detail.

As already mentioned above, the control of cell cycle progression in eukaryotes is mainly exerted at two transition points: one in late G_1, before DNA synthesis, and one at the G_2/M boundary. Progression through these control points is mediated by cyclin-dependent protein kinase (CDK) complexes, which contain, in more detail, a catalytic subunit of approximately 34-kDa encoded by the *CDK* genes. Both *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* only utilize one *CDK* gene for the regulation of their cell cycle. The kinase activity of their gene products $p34^{CDC2}$ and $p34^{CDC28}$ in *Sch. pombe* and in *S. cerevisiae*, respectively, is dependent on regulatory proteins, called cyclins. Progression through the different cell cycle phases is achieved by the sequential association of $p34^{CDC2/CDC28}$ with different cyclins. Although in higher eukaryotes this regulation mechanism is conserved, the situation is more complex since

they have evolved to use multiple CDKs to regulate the different stages of the cell cycle. In mammals, seven CDKs have been described, defined as CDK1 to CDK7, each binding a specific subset of cyclins.

In animal systems, CDK activity is not only regulated by its association with cyclins but also involves both stimulatory and inhibitory phosphorylations. Kinase activity is positively regulated by phosphorylation of a Thr residue located between amino acids 160-170 (depending on the CDK protein). This phosphorylation is mediated by the CDK-activating kinase (CAK) which interestingly is a CDK/cyclin complex itself. Inhibitory phosphorylations occur at the ATP-binding site (the Tyr15 residue together with Thr14 in higher eukaryotes) and are carried out by at least two protein kinases. A specific phosphatase, CDC25, dephosphorylates these residues at the G_2/M checkpoint, thus activating CDK activity and resulting in the onset of mitosis. CDK activity is furthermore negatively regulated by a family of mainly low-molecular weight proteins, called cyclin-dependent kinase inhibitors (CKIs). Kinase activity is inhibited by the tight association of these CKIs with the CDK/cyclin complexes.

With respect to cell cycle regulation in plants a summary of the state of the art is given below. In *Arabidopsis*, thusfar only two *CDK* genes have been characterized in detail, *CDC2aAt* and *CDC2bAt*, of which the gene products share 56% amino acid identity. Both CDKs are distinguished by several features. First, only *CDC2aAt* is able to complement yeast $p34^{CDC2/CDC28}$ mutants. Second, CDC2aAt and CDC2bAt bear different cyclin-binding motifs (PSTAIRE and PPTALRE, respectively), suggesting they may bind distinct types of cyclins. Third, although both *CDC2aAt* and *CDC2bAt* show the same spatial expression pattern, they exhibit a different cell cycle phase-specific regulation. The *CDC2aAt* gene is expressed constitutively throughout the whole cell cycle. In contrast, *CDC2bAt* mRNA levels oscillate, being most abundant during the S and G_2 phases. In addition, multiple cyclins have been isolated from *Arabidopsis*. The majority displays the strongest sequence similarity with the animal A- or B-type class of cyclins, but also D-type cyclins have been identified. Although the classification of *Arabidopsis* cyclins is mainly based upon sequence similarity, limited data suggests that this organization corresponds with differential functions of each cyclin class (Renaudin, Plant Mol. Biol. 32 (1996) 1003-1018).

In order to manage problems related to plant growth, plant architecture, stress responses and/or plant diseases, it is believed to be of utmost importance to identify and isolate plant genes and gene products involved in the regulation of the plant cell division, and more particularly coding for and interacting with CDKs and/or their interacting proteins, responsible for the control of the cell cycle. If such novel genes and/or proteins have been isolated and analyzed; the growth of the plant as a whole can be influenced. Also, the growth of specific tissues or organs and thus the architecture of the plant can be modified. Cell cycle proteins may also provide targets to facilitate the identification of inhibitors or activators of cell cycle regulatory proteins that may useful as herbicides or plant growth regulators.

Thus, the technical problem underlying the present invention is to provide means and methods for modulating cell cycle proteins that are particular useful in agriculture and plant cell and tissue culture.

The solution to the technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to a DNA sequence encoding a cell cycle interacting protein or encoding an immunologically active and/or functional fragment of such a protein, selected from the group consisting of:

(a) DNA sequences

(aa) comprising a nucleotide sequence encoding at least the mature form of a protein (LDV115) comprising the amino acid sequence as given in SEQ ID NO: 2;

(ab) comprising the nucleotide sequence as given in SEQ ID NO: 1;

(ac) comprising a nucleotide sequence hybridizing with the complementary strand of a nucleotide sequence as defined in (aa) or (ab) under stringent hybridization conditions;

(ad) comprising an nucleotide sequence encoding a protein having an amino acid sequence at least 60 % identical to the amino acid sequence encoded by the nucleotide sequence of (aa) or (ab);

(ae) comprising a nucleotide sequence encoding at least the domain binding to CDKs of the protein encoded by the nucleotide sequence of any one of (aa) to (ad);

(b) DNA sequences

(ba) comprising a nucleotide sequence encoding at least the mature form of a PHO80-like Protein (PLP) comprising the amino acid sequence as given in any one of SEQ ID NOs: 4, 34, 36, 38, 40 or 42;

(bb) comprising the nucleotide sequence as given in any one of SEQ ID NOs: 3, 33, 35, 37, 39 or 41;

(bc) comprising a nucleotide sequence hybridizing with the complementary strand of a nucleotide sequence as defined in (ba) or (bb) under stringent hybridization conditions;

(bd) comprising an nucleotide sequence encoding a protein having an amino acid sequence at least 40 % identical to the amino acid sequence encoded by the nucleotide sequence of (ba) or (bb);

(be) comprising a nucleotide sequence encoding at least the cyclin-like interacting domain of the protein encoded by the nucleotide sequence of any one of (ba) to (bd);

(c) DNA sequences

(ca) comprising a nucleotide sequence encoding at least the mature form of a protein (VB33) comprising the amino acid sequence as given in SEQ ID NO: 6;

(cb) comprising the nucleotide sequence as given in SEQ ID NO: 5;

(cc) comprising a nucleotide sequence hybridizing with the complementary strand of a nucleotide sequence as defined in (ca) or (cb) under stringent hybridization conditions;

(cd) comprising an nucleotide sequence encoding a protein having an amino acid sequence at least 60 % identical to the amino acid sequence encoded by the nucleotide sequence of (ca) or (cb);

(ce) comprising a nucleotide sequence encoding at least the domain binding to CDKs of the protein encoded by the nucleotide sequence of any one of (ca) to (cd);

(d) DNA sequences

(da) comprising a nucleotide sequence encoding at least the mature form of a protein (VB89) comprising the amino acid sequence as given in SEQ ID NO: 8;

(db) comprising the nucleotide sequence as given in SEQ ID NO: 7;

(dc) comprising a nucleotide sequence hybridizing with the complementary strand of a nucleotide sequence as defined in (da) or (db) under stringent hybridization conditions;

(dd) comprising an nucleotide sequence encoding a protein having an amino acid sequence at least 60 % identical to the amino acid sequence encoded by the nucleotide sequence of (da) or (db);

(de) comprising a nucleotide sequence encoding at least the domain binding to CDKs of the protein encoded by the nucleotide sequence of any one of (da) to (dd);

(e) DNA sequences

(ea) comprising a nucleotide sequence encoding at least the mature form of a protein (VBDAHP) comprising the amino acid sequence as given in SEQ ID NO: 10;

(eb) comprising the nucleotide sequence as given in SEQ ID NO: 9;

(ec) comprising a nucleotide sequence hybridizing with the complementary strand of a nucleotide sequence as defined in (ea) or (eb) under stringent hybridization conditions;

(ed) comprising an nucleotide sequence encoding a protein having an amino acid sequence at least 60 % identical to the amino acid sequence encoded by the nucleotide sequence of (ea) or (eb);

(ee) comprising a nucleotide sequence encoding at least the domain binding to CDKs of the protein encoded by the nucleotide sequence of any one of (ea) to (ed);

(f) DNA sequences

(fa) comprising a nucleotide sequence encoding at least the mature form of a protein (VBDBP) comprising the amino acid sequence as given in SEQ ID NO: 12;

(fb) comprising the nucleotide sequence as given in SEQ ID NO: 11;

(fc) comprising a nucleotide sequence hybridizing with the complementary strand of a nucleotide sequence as defined in (fa) or (fb) under stringent hybridization conditions;

(fd) comprising an nucleotide sequence encoding a protein having an amino acid sequence at least 60 % identical to the amino acid sequence encoded by the nucleotide sequence of (fa) or (fb);

(fe) comprising a nucleotide sequence encoding at least the domain binding to CDKs of the protein encoded by the nucleotide sequence of any one of (fa) to (fd);

(g) DNA sequences

(ga) comprising a nucleotide sequence encoding at least the mature form of a protein (VBHSF) comprising the amino acid sequence as given in SEQ ID NO: 14;

(gb) comprising the nucleotide sequence as given in SEQ ID NO: 13;

(gc) comprising a nucleotide sequence hybridizing with the complementary strand of a nucleotide sequence as defined in (ga) or (gb) under stringent hybridization conditions;

(gd) comprising an nucleotide sequence encoding a protein having an amino acid sequence at least 60 % identical to the amino acid sequence encoded by the nucleotide sequence of (ga) or (gb);

(ge) comprising a nucleotide sequence encoding at least the domain binding to CDKs of the protein encoded by the nucleotide sequence of any one of (ga) to (gd);

(h) DNA sequences obtainable by screening an appropriate library under stringent conditions with a probe having at least 17 consecutive nucleotides of a nucleotide sequence of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15 to 33, 35, 37, 39, 41, 48, 49 or 53 to 57;

(i) DNA sequences comprising a nucleotide sequence encoding a fragment of a protein encoded by a DNA sequence of any one of (a) to (h), wherein said fragment is capable of interacting with a cell cycle protein; and

(j) DNA sequences, the nucleotide sequence of which is degenerate as a result of the genetic code to a nucleotide sequence of a DNA sequence as defined in any one of (a) to (i).

The term "cell cycle interacting protein" or "cell cycle protein" as denoted herein means a protein which exerts control on or regulates or is required for the cell cycle or part thereof of a cell, tissue, organ or whole organism and/or DNA replication. It is may also be capable of binding to, regulating or being regulated by cyclin dependent kinases, in particular CDC2a and/or CDC2b and preferably to plant cyclin dependent kinases or their subunits. The term also includes peptides, polypeptides, fragments, variant, homologs, alleles or precursors (eg preproteins or proproteins) thereof.

The term "cell cycle" means the cyclic biochemical and structural events associated with growth, division and proliferation of cells, and in particular with the regulation of the replication of DNA and mitosis. The cycle is divided into periods called: G_0, Gap_1 (G_1), DNA synthesis (S), Gap_2 (G_2), and mitosis (M). Normally these four phases occur sequentially, however the cell cycle also includes modified cycles wherein one or more phases are absent resulting in modified cell cycle such as endomitosis, acytokinesis, polyploidy, polyteny, and endoreduplication.

The terms "gene", "polynucleotide", "nucleic acid sequence", "nucleotide sequence", "DNA sequence" or "nucleic acid molecule" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA, and RNA. It also includes known types of modifications, for example, methylation, "caps" substitution of one or more of the naturally occuring nucleotides with an analog. Preferably, the DNA sequence of the invention comprises a coding sequence encoding at least the mature form of the above defined cell cycle interacting protein, i.e. the protein which is posttranslationally processed in its biologically active form, for example due to cleavage of leader or secretory sequences or a proprotein sequence or other natural proteolytic cleavage points.

By "functional fragment" and "biologically active form" polypeptides are meant that exhibit activity similar, but not necessarily identical, to an activity of the wild-type cell cycle interacting proteins of the invention or an activity that is enhanced over that of the

wild-type proteins (either the full-length protein or, preferably, the mature protein), as measured in a particular biological assay. Assays of cell cycle interacting activity are disclosed, for example, in Examples 1 to 7, below. These assays can be used to measure cell cycle interacting activity of partially purified or purified native or recombinant protein. The cell cycle interacting protein of the invention binds to CDC2, i.e. CDC2a and/or CDC2b, e.g., from Arabidopsis. Thus, a polypeptide having a functional fragment or the "biological activity" of the cell cycle interacting protein of the invention will bind to CDCs as set forth in Example 1 or 7.

The term "immunologically active fragment" of a cell cycle interacting protein of the invention denotes proteins or peptides which have at least a part of the primary structural conformation for one or more epitopes capable of reacting specifically with antibodies to a protein which is encodable by a nucleic acid molecule as set forth above. Preferably, the peptides and proteins encoded by a nucleic acid molecule of the invention are recognized by an antibody that specifically recognizes an epitope of the cell cycle interacting protein comprising the amino acid residues that are unique for the protein encoded by any one of the aforementioned DNA sequences. Preferably, said peptides and proteins are capable of eliciting an effective immune response in a mammal, for example mouse or rabbit.

The DNA sequence which encodes for the predicted mature polypeptides of the proteins comprising SEQ ID NOS: 2, 4, 34, 36, 38, 40, 42, 6, 8, 10, 12 or 14 or for the biologically active fragment thereof may include: only the coding sequence for the mature polypeptide or for a biologically active fragment thereof; the coding sequence for the mature polypeptide or for a biologically active fragment thereof and additional coding sequence such as a leader or secretory sequence or a proprotein sequence; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as intron or non-coding sequence 5' and/or 3' of the coding sequence for the predicted mature polypeptide.

A "coding sequence" is a nucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, recombinant nucleotide

sequences or genomic DNA, while introns may be present as well under certain circumstances. Thus, the nucleotide sequences of the present invention can be engineered in order to alter a cell cycle interacting protein coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, e.g., site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.

In accordance with the present invention a two-hybrid system (Fields et al., Nature 340 (1989), 245-246,) was exploited whereby CDC2aAt or CDC2bAt as bait and a cDNA library of a cell suspension as prey are used. Novel gene products interacting with CDC2aAt or CDC2bAt indicative of hitherto unknown plant cell cycle regulatory nucleotide sequences were identified. The library was made from a mixture mRNA from Arabidopsis thaliana cell suspensions harvested at various growing stages: early exponential, exponential, early stationary and stationary phase.

Twelve cDNA clones have been identified in accordance with the invention comprising the nucleotide sequences as depicted in SEQ ID NOS: 1, 3, 33, 35, 37, 39, 41, 5, 7, 9, 11 and 13, which encode proteins that are capable of specifically interacting with cdc2aAt or cdc2bAt; see Examples 1, 2 and 7, below. The proteins encoded by the cDNA clones comprised the amino acid sequences depicted in SEQ ID NOS: 2, 4, 34, 36, 38, 40, 42, 6, 8, 10, 12 and 14. Computer assisted homology search in genome data bases revealed that novel genes have been identified and/or genes where the (partial) cDNA was described but the particular function of the gene remained unknown. In particular, the examples of the present invention demonstrate that novel cell cycle interacting proteins and their encoding genes have been identified. The possible applications of the these cell cycle interacting proteins and their encoding nucleic acid molecules will be discussed further below and are evident from the description provided in the Examples.

The homology search was performed with the program BLASTX and BLASTN (version 2.0a19MP-WashU [build decunix3.2 01:53:29 05-feb-1998] (see Altschul, Nucleic Acids Res. 25 (1997), 3389-3402) on the Arabidopsis thaliana nucleic acids database at ATDB at Stanford (http://genome-www2.stanford.edu/cgi-bin/AtDB/nph-blast2atdb). The function GAP (general alignment) (from the GCG 9.1 package, Genetics Computer Group Inc., Madison, USA) has been used with the parameters Gap weight = 12 and Length weight = 4 to quantify the percentage of homology and similarity. The protein sequences were then used to perform a BLASTP (version 2.0.4 [feb-24-1998]) with BEAUTY post-processing provided by the Human Genome Center, Baylor College of Medicine against the National Center for Biotechnology Information's non-redundant protein database (http://dot.imgen.bcm.tmc.edu:9331/seq-search/protein-search.html). The results of the homology search are described in the appended examples.

As described in the Examples, during the course of search in the database homology has been found for one or more of the above described nucleotide sequences to some "Expressed Sequence Tags" (ESTs), i.e. (partial) cDNA clones comprising Open Reading Frames (ORFs) for (fragments of) proteins of unknown function and/or the nucleotide sequence of which has not sufficient coding capacity for a functional protein. These particular ESTs per se are specifically excluded from the scope of the claims. However, as far as the use of such ESTs in embodiments is concerned which have been first conceived in accordance with the present invention they are covered by the present invention and encompassed by the appended claims. The same applies to nucleotide sequences that may be present within for example a section of a chromomsome that has been described in context with an organism's genome sequencing project but hitherto have not been identified to constitute a gene with biological function, nor what the particular biological function of this gene could be.

Thus it is evident that the genes comprising the nucleotide sequences of each SEQ ID NOS: 1, 3, 33, 35, 37, 39, 41, 5, 7, 9, 11 and 13 each encode a member of a novel class of cell cycle interacting proteins. In particular, the nucleotide sequences of SEQ ID NOS: 3, 33, 35, 37, 39 and 41 define a novel class of PHO80-like Proteins (PLPs); see also Example 7.

The present invention also relates to DNA sequences hybridizing with the above-described DNA sequences and differ in one or more positions in comparison with these as long as they encode a cell cycle interacting protein. By "hybridizing" it is meant that such nucleic acid molecules hybridize under conventional hybridization conditions, preferably under stringent conditions such as described by, e.g., Sambrook (Molecular Cloning; A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)). An example of one such stringent hybridization condition is hybridization at 4XSSC at 65 ºC, followed by a washing in 0.1XSSC at 65 ºC for one hour. Alternatively, an exemplary stringent hybridization condition is in 50 % formamide, 4XSSC at 42 ºC. Cell cycle interacting proteins derived from other organisms such as mammals, in particular humans, may be encoded by other DNA sequences which hybridize to the sequences for plant cell cycle interacting proteins under relaxed hybridization conditions and which code on expression for peptides having the ability to interact with cell cycle proteins. Examples of such non-stringent hybridization conditions are 4XSSC at 50 ºC or hybridization with 30-40 % formamide at 42 ºC. Further preferred hybridization conditions are described in the examples. Such molecules comprise those which are fragments, analogues or derivatives of the cell cycle interacting protein of the invention and differ, for example, by way of amino acid and/or nucleotide deletion(s), insertion(s), substitution(s), addition(s) and/or recombination(s) or any other modification(s) known in the art either alone or in combination from the above-described amino acid sequences or their underlying nucleotide sequence(s). Using the PESTFIND program (Rogers, Science 234 (1986), 364-368), PEST sequences (rich in proline, glutamic acid, serine, and threonine) can be identified, which are characteristically present in unstable proteins. Such sequences may be removed from the cell cycle interacting proteins in order to increase the stability and optionally the activity of the proteins. Methods for introducing such modifications in the nucleic acid molecules according to the invention are well-known to the person skilled in the art. The invention also relates to nucleic acid molecules the sequence of which differs from the nucleotide sequence of any of the above-described nucleic acid molecules due to the degeneracy of the genetic code. All such fragments, analogues and derivatives of the protein of the invention are included within the scope of the present invention, as long as the essential characteristic immunological and/or biological properties as defined above remain unaffected in kind,

that is the novel nucleic acid molecules of the invention include all nucleotide sequences encoding proteins or peptides which have at least a part of the primary structural conformation for one or more epitopes capable of reacting with antibodies to cell cycle interacting proteins which are encodable by a nucleic acid molecule as set forth above and which have comparable or identical characteristics in terms of biological activity and/or the capability to interact with other proteins. It is preferred that proteins encoded by a nucleic acid molecule of the invention are at least capable of interacting with CDC2, particularly CDC2a and/or CDC2b, preferably from a plant such as Arabidopsis thaliana. Whilst the above described proteins may interact with a CDC2 from *Arabidopsis thaliana*, the most likely interaction is with a CDC2 from the same species from which the gene was isolated (homologous interaction). This capability allows advantageous uses of the proteins of the invention and their encoding nucleic acid molecules as will be described in more detail below. Part of the invention is therefore also nucleic acid molecules encoding a polypeptide comprising at least a functional part of a cell cycle interacting protein encoded by a nucleic acid sequence comprised in a nucleic acid molecule according to the invention. An example for this is that the polypeptide or a fragment thereof according to the invention is embedded in another amino acid sequence. Preferably, the DNA sequence of the invention encodes a protein having substantially the same amino acid sequence as the proteins defined in SEQ ID NOS: 2, 4, 34, 36, 38, 40, 42, 6, 8, 10, 12 and 14.

Extending the polynucleotide sequence of the invention

The polynucleotide sequences encoding the cell cycle interacting proteins may be extended utilizing partial nucleotide sequence and various methods known in the art to detect upstream sequences such as promoters and regulatory elements. Gobinda, (PCR Methods Applic. 2 (1993), 318-322) discloses "restriction-site" polymerase chain reaction (PCR) as a direct method which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are

transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia, Nucleic Acids Res. 16 (1988), 8186). The primers may be designed using OLIGO® 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth MN), or another appropriate program to be preferably 22-30 nucleotides in length, to have a GC content of preferably 50% or more, and to anneal to the target sequence at temperatures preferably about 68°-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom, PCR Methods Applic. 1 (1991), 111-119) is a method for PCR amplification of DNA fragments adjacent to a known sequence in, e.g., human or plant yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, (Nucleic Acids Res. 19 (1991), 3055-3060). Additionally, one can use PCR, nested primers and PromoterFinder libraries to walk in genomic DNA (PromoterFinder™ Clontech (Palo Alto CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region. Suitable methods for identifying promoters are also described in WO 99/61619, in particular at pages 50 and 51. Capillary electrophoresis may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products; see, e.g., Sambrook, supra. Systems for rapid sequencing are available from Perkin Elmer, Beckmann Instruments (Fullerton CA), and other companies.

Computer-assisted identification of cell cycle interacting proteins and their encoding genes

As is further described in the appended examples BLAST2, which stands for Basic Local Alignment Search Tool (Altschul, 1997; Altschul, J. Mol. Evol. 36 (1993), 290-300; Altschul, J. Mol. Biol. 215 (1990), 403-410), can be used to search for local sequence alignments. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP). An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

Analogous computer techniques using BLAST (Altschul, 1997, 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or EMBL. This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous. The basis of the search is the product score which is defined as:

$$\frac{\%\text{sequence identity} \times \%\text{ maximum BLAST score}}{100}$$

and it takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules

are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

Identifying derivatives, variants and homologs of the cell cycle interacting proteins of the invention

As is demonstrated in the appended examples a two-hybrid screening assay has been developed in accordance with the present invention suitable for identifying cell cycle interacting proteins. Thus, in another aspect the present invention relates to a method for identifying and obtaining cell cycle interacting proteins comprising a two-hybrid screening assay wherein CDC2a or CDC2b as a bait and a cDNA library of cell suspension culture as prey are used. Preferably, said CDC2a and CDC2b is CDC2aAt and CDC2bAt, respectively. However, CDKs or their corresponding subunits from other plants or other organisms such as mammals may be employed as well. The cell culture may be from any organism possessing cell cycle interacting proteins such as animals, preferably mammals. Particularly preferred are plant cell suspension cultures such as from Arabidopsis. The nucleic acid molecules encoding proteins or peptides identified to interact with CDC2a or CDC2b in the above mentioned assay can be easily obtained and sequenced by methods known in the art; see also the appended examples. Therefore, the present invention also relates to a DNA sequence encoding a cell cycle interacting protein obtainable by the method of the invention.

In a preferred embodiment the nucleic acid molecules according to the invention are RNA or DNA molecules, preferably cDNA, genomic DNA or synthetically synthesized DNA or RNA molecules. Since cell cycle interacting proteins are supposed to play a key role in the plant cell cycle, corresponding proteins displaying similar properties should be present in other organisms including mammals as well. Nucleic acid molecules of the invention can be obtained, e.g., by hybridization of the above-described nucleic acid molecules with a (sample of) nucleic acid molecule(s) of any source. Nucleic acid molecules hybridizing with the above-described nucleic acid molecules can in general be derived from any organism, preferably plants possessing such molecules, preferably from monocotyledonous or dicotyledonous plants, in particular from plants of interest in

agriculture, horticulture or wood culture, such as crop plants, namely those of the family Poaceae, any starch producing plants, such as potato, maniok, leguminous plants, oil producing plants, such as oilseed rape, linenseed, etc., plants using polypeptide as storage substances, such as soybean, plants using sucrose as storage substance, such as sugar beet or sugar cane, trees, ornamental plants etc. Preferably, the nucleic acid molecules according to the invention are derived from crop plants (e.g. maize, rice, barley, wheat, rye, oats etc.), potatoes, oil producing plants (e.g. oilseed rape, sunflower, peanut, soybean, etc.), cotton, sugar beet, sugar cane, leguminous plants (e.g. beans, peas etc.), and, of course, from *Arabidopsis thaliana*. Nucleic acid molecules hybridizing to the above-described nucleic acid molecules can be isolated, e.g., from libraries, such as cDNA or genomic libraries by techniques well known in the art. For example, hybridizing nucleic acid molecules can be identified and isolated by using the above-described nucleic acid molecules or fragments thereof or complements thereof as probes to screen libraries by hybridizing with said molecules according to standard techniques. Possible is also the isolation of such nucleic acid molecules by applying a nucleic acid amplicification technique such as the polymerase chain reaction (PCR) using as primers oligonucleotides derived form the above-described nucleic acid molecules. Also nucleic acid molecules may be identified and isolated using microarrays or DNA chips (Southern et al. (1999) Nat. Genet, Jan:21(1 Suppl.):5-9; Ramsay, (1998) Nature Biotechnology, 16 (1):40).

Nucleic acid molecules which hybridize with any of the aforementioned nucleic acid molecules also include fragments, derivatives and allelic variants of the above-described nucleic acid molecules that encode a cell cycle interacting protein or an immunologically active or functional fragment thereof. Fragments are understood to be parts of nucleic acid molecules long enough to encode the described protein or a functional or immunologically active fragment thereof as defined above.

The term "derivative" means in this context that the nucleotide sequence of these nucleic acid molecules differs from the sequences of the above-described nucleic acid molecules in one or more nucleotide positions and are highly homologous to said nucleic acid molecules. Homology is understood to refer to a sequence identity of at least 40 %, particularly an identity of at least 60 %, preferably more than 80 % and still more

preferably more than 90 %. The term "substantially homologous" refers to a subject, for instance a nucleic acid, which is at least 50% identical in sequence to the reference when the entire ORF (open reading frame) is compared, where the sequence identity is preferably at least 70%, more preferably at least 80%, still more preferably at least 85%, especially more than about 90%, most preferably 95% or greater, particularly 98% or greater. The deviations from the sequences of the nucleic acid molecules described above can, for example, be the result of nucleotide substitution(s), deletion(s), addition(s), insertion(s) and/or recombination(s); see supra.

Homology further means that the respective nucleic acid molecules or encoded proteins may also be functionally and/or structurally equivalent. The nucleic acid molecules that are homologous to the nucleic acid molecules described above and that are derivatives of said nucleic acid molecules are, for example, variations of said nucleic acid molecules which represent modifications having the same biological function, in particular encoding proteins with the same or substantially the same biological function. They may be naturally occurring variations, such as sequences from other plant varieties or species, or mutations. These mutations may occur naturally or may be obtained by mutagenesis techniques. The allelic variations may be naturally occurring allelic variants as well as synthetically produced or genetically engineered variants; see supra.

The proteins encoded by the various derivatives and variants of the above-described nucleic acid molecules may share specific common characteristics, such as biological activity, molecular weight, immunological reactivity, conformation, etc., as well as physical properties, such as electrophoretic mobility, chromatographic behavior, sedimentation coefficients, pH optimum, temperature optimum, stability, solubility, spectroscopic properties, etc.

Examples of the different possible applications of the nucleic acid molecules according to the invention as well as molecules derived from them will be described in detail in the following.

Uses of the nucleic acid molecules of the present invention

In one embodiment, the present invention relates to a nucleic acid molecule which hybridizes with the complementary strand of the nucleic acid molecule of the invention and which encodes a mutated version of the protein as defined above which has lost its immunological and/or biological activity. This embodiment may prove useful for, e.g., generating dominant mutant alleles of the above-described cell cycle interacting proteins. Said mutated version is preferably generated by substitution, deletion and/or addition of 1 to 5 or 5 to 10 amino acid residues in the amino acid sequence of the above-described wild type proteins.

In a further embodiment, the invention relates to nucleic acid molecules of at least 15 nucleotides in length hybridizing specifically with a nucleic acid molecule as described above or with a complementary strand thereof. Specific hybridization occurs preferably under stringent conditions and implies no or very little cross-hybridization with nucleotide sequences encoding no or substantially different proteins. Such nucleic acid molecules may be used as probes and/or for the control of gene expression. Nucleic acid probe technology is well known to those skilled in the art who will readily appreciate that such probes may vary in length. Preferred are nucleic acid probes of 16 to 35 nucleotides in length. Of course, it may also be appropriate to use nucleic acids of up to 100 and more nucleotides in length. The nucleic acid probes of the invention are useful for various applications. On the one hand, they may be used as primers for amplification of nucleic acid sequences according to the invention. The design and use of said primers is known by the person skilled in the art. Preferably such amplification primers comprise a contiguous sequence of at least 6 nucleotides, in particular 13 nucleotides, preferably 15 to 25 nucleotides or more, identical or complementary to the nucleotide sequence depicted in SEQ ID NOS: 1, 3, 33, 35, 37, 39, 41, 5, 7, 9, 11 or 13 or to a nucleotide sequence encoding the amino acid sequence of SEQ ID NOS: 2, 4, 34, 36, 38, 40, 42, 6, 8, 10, 12 or 14. Another application is the use as a hybridization probe to identify nucleic acid molecules hybridizing with a nucleic acid molecule of the invention by homology screening of genomic DNA or cDNA libraries. Nucleic acid molecules according to this preferred embodiment of the invention which are complementary to a

nucleic acid molecule as described above are preferably at least 17 nucleotides in length and may also be used for repression of expression of a cell cycle gene, for example due to an antisense or triple helix effect or for the construction of appropriate ribozymes (see, e.g., EP-A1 0 291 533, EP-A1 0 321 201, EP-A2 0 360 257) which specifically cleave the (pre)-mRNA of a gene comprising a nucleic acid molecule of the invention or part thereof. Selection of appropriate target sites and corresponding ribozymes can be done as described, for example, in Steinecke, Ribozymes, Methods in Cell Biology 50, Galbraith et al.. eds Academic Press, Inc. (1995), 449-460. The above described nucleic acid molecules may either be DNA or RNA or a hybrid thereof. Furthermore, said nucleic acid molecule may contain, for example, thioester bonds and/or nucleotide analogues, commonly used in oligonucleotide anti-sense approaches. Said modifications may be useful for the stabilization of the nucleic acid molecule against endo- and/or exonucleases in the cell. Said nucleic acid molecules may be transcribed by an appropriate vector containing a chimeric gene which allows for the transcription of said nucleic acid molecule in the cell.

Furthermore, the person skilled in the art is well aware that it is also possible to label such a nucleic acid probe with an appropriate marker for specific applications, such as for the detection of the presence of a nucleic acid molecule of the invention in a sample derived from an organism, in particular plants. A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US Patents US-A-3,817,837; US-A-3,850,752; US-A-3,939,350; US-A-3,996,345; US-A-4,227,437; US-A-4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced as shown in US-A-4,816,567 incorporated herein by reference.

Furthermore, the so-called "peptide nucleic acid" (PNA) technique can be used for the detection or inhibition of the expression of a nucleic acid molecule of the invention. For example, the binding of PNAs to complementary as well as various single stranded RNA and DNA nucleic acid molecules can be systematically investigated using thermal denaturation and BIAcore surface-interaction techniques (Jensen, Biochemistry 36

(1997), 5072-5077). Furthermore, the nucleic acid molecules described above as well as PNAs derived therefrom can be used for detecting point mutations by hybridization with nucleic acids obtained from a sample with an affinity sensor, such as BIAcore; see Gotoh, Rinsho Byori 45 (1997), 224-228. Hybridization based DNA screening on peptide nucleic acids (PNA) oligomer arrays are described in the prior art, for example in Weiler, Nucleic Acids Research 25 (1997), 2792-2799. The synthesis of PNAs can be performed according to methods known in the art, for example, as described in Koch, J. Pept. Res. 49 (1997), 80-88; Finn, Nucleic Acids Research 24 (1996), 3357-3363. Further possible applications of such PNAs, for example as restriction enzymes or as templates for the synthesis of nucleic acid oligonucleotides are known to the person skilled in the art and are, for example, described in Veselkov, Nature 379 (1996), 214 and Bohler, Nature 376 (1995), 578-581. A further application of the nucleic acids of the invention is their use in a two-hybrid system to identify interacting proteins (i.e. proteins that specifically interact with the nucleic acid-encoding products). Methods for preparing and performing the two-hybrid screen are known in the art, including descriptions provided in this document and generally see Hannon G. and Bartel P. Identification of interacting proteins using the two-hybrid system Methods Mol. Cellular Biol. 5 (1995), 289-297.

Detection and mapping of related polynucleotide sequences

The nucleic acid sequence for a cell cycle interacting protein of the invention can also be used to generate hybridization probes for mapping the naturally occurring genomic sequence. The sequence may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include *in situ* hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price (Blood Rev. 7 (1993), 127-134) and Trask (Trends Genet. 7 (1991), 149-154). The technique of fluorescent *in situ* hybridization of chromosome spreads has been described, among other places, in Verma, (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY. Fluorescent *in situ*

hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f) and Meinke, Science 282 (1998), 662-682. Correlation between the location of the gene encoding a cell cycle interacting protein of the invention on a physical chromosomal map and a specific feature, e.g., plant growth, architecture, yield, stress, disease etc. may help delimit the region of DNA associated with this feature. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals. Furthermore, the means and methods described herein can be used for marker-assisted breeding.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. For example an sequence tagged site based map of the human genome was recently published by the Whitehead-MIT Center for Genomic Research (Hudson, Science 270 (1995), 1945-1954) on a map of the plant genome by way of the Arabidopsis genome is available from http://genome.wwz.Stanford.edu/cgi-bin/AtDB/nph-blast2atdb. Often the placement of a gene on the chromosome of another species may reveal associated marker even if the number or arm of a particular chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for interacting genes using positional cloning or other gene discovery techniques. Once such gene has been crudely localized by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals.

Vectors and expression systems

The present invention also relates to vectors, particularly plasmids, cosmids, viruses, bacteriophages and other vectors used conventionally in genetic engineering that

contain a nucleic acid molecule according to the invention. Methods which are well known to those skilled in the art can be used to construct various plasmids and vectors; see, for example, the techniques described in Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. (1989), (1994). Plasmids and vectors to be preferably employed in accordance with the present invention include those well known in the art. Alternatively, the nucleic acid molecules and vectors of the invention can be reconstituted into liposomes for delivery to target cells.

In a preferred embodiment the nucleic acid molecule present in the vector is linked to (a) control sequence(s) which allow the expression of the nucleic acid molecule in prokaryotic and/or eukaryotic cells.

The term "control sequence" refers to regulatory DNA sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism. In prokaryotes, control sequences generally include promoter, ribosomal binding site, and terminators. In eukaryotes generally control sequences include promoters, terminators and, in some instances, enhancers, transactivators or transcription factors. The term "control sequence" is intended to include, at a minimum, all components the presence of which are necessary for expression, and may also include additional advantageous components.

The term "operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. In case the control sequence is a promoter, it is obvious for a skilled person that double-stranded nucleic acid is preferably used.

Thus, the vector of the invention is preferably an expression vector. An "expression vector" is a construct that can be used to transform a selected host cell and provides for expression of a coding sequence in the selected host. Expression vectors can for instance be cloning vectors, binary vectors or integrating vectors. Expression comprises

transcription of the nucleic acid molecule preferably into a translatable mRNA. Regulatory elements ensuring expression in prokaryotic and/or eukaryotic cells are well known to those skilled in the art. In the case of eukaryotic cells they comprise normally promoters ensuring initiation of transcription and optionally poly-A signals ensuring termination of transcription and stabilization of the transcript, for example in plants, those of the 35S RNA from Cauliflower Mosaic Virus (CaMV). Other promoters commonly used are the polyubiquitin promoter, and the actin promoter for ubiquitous expression. The termination signals usually employed are from the Nopaline Synthase promoter or from the CAMV 35S promoter. A plant translational enhancer often used is the TMV omega sequences, the inclusion of an intron (Intron-1 from the Shrunken gene of maize, for example) has been shown to increase expression levels by up to 100-fold. (Mait, Transgenic Research 6 (1997), 143-156; Ni, Plant Journal 7 (1995), 661-676). Additional regulatory elements may include transcriptional as well as translational enhancers. Possible regulatory elements permitting expression in prokaryotic host cells comprise, e.g., the P_L, *lac*, *trp* or *tac* promoter in *E. coli*, and examples of regulatory elements permitting expression in eukaryotic host cells are the *AOX1* or *GAL1* promoter in yeast or the CMV-, SV40-, RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. In this context, suitable expression vectors are known in the art such as Okayama-Berg cDNA expression vector pcDV1 (Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3 (In-vitrogene), pSPORT1 (GIBCO BRL). An alternative expression system which could be used to express a cell cycle interacting protein is an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The coding sequence of a nucleic acid molecule of the invention may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of said coding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frugiperda* cells or *Trichoplusia* larvae in which the protein of the invention is expressed (Smith, J. Virol. 46 (1983), 584; Engelhard, Proc. Nat. Acad. Sci. USA 91 (1994), 3224-3227). Further promoters and expression systems that may be

used in accordance with the present invention are described in the prior art, for example WO 99/61619.

Advantageously, the above-described vectors of the invention comprises a selectable and/or scorable marker. Selectable marker genes useful for the selection of transformed plant cells, callus, plant tissue and plants are well known to those skilled in the art and comprise, for example, antimetabolite resistance as the basis of selection for dhfr, which confers resistance to methotrexate (Reiss, Plant Physiol. (Life Sci. Adv.) 13 (1994), 143-149); npt, which confers resistance to the aminoglycosides neomycin, kanamycin and paromycin (Herrera-Estrella, EMBO J. 2 (1983), 987-995) and hygro, which confers resistance to hygromycin (Marsh, Gene 32 (1984), 481-485). Additional selectable genes have been described, namely trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman, Proc. Natl. Acad. Sci. USA 85 (1988), 8047); mannose-6-phosphate isomerase which allows cells to utilize mannose (WO 94/20627) and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue, 1987, In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory ed.) or deaminase from Aspergillus terreus which confers resistance to Blasticidin S (Tamura, Biosci. Biotechnol. Biochem. 59 (1995), 2336-2338).

Useful scorable marker are also known to those skilled in the art and are commercially available. Advantageously, said marker is a gene encoding luciferase (Giacomin, Pl. Sci. 116 (1996), 59-72; Scikantha, J. Bact. 178 (1996), 121), green fluorescent protein (Gerdes, FEBS Lett. 389 (1996), 44-47) or ß-glucuronidase (Jefferson, EMBO J. 6 (1987), 3901-3907). This embodiment is particularly useful for simple and rapid screening of cells, tissues and organisms containing a vector of the invention.

The present invention furthermore relates to host cells comprising a vector as described above or a nucleic acid molecule according to the invention wherein the nucleic acid molecule is foreign to the host cell.

By "foreign" it is meant that the nucleic acid molecule is either heterologous with respect to the host cell, this means derived from a cell or organism with a different genomic background, or is homologous with respect to the host cell but located in a different

genomic environment than the naturally occurring counterpart of said nucleic acid molecule. This means that, if the nucleic acid molecule is homologous with respect to the host cell, it is not located in its natural location in the genome of said host cell, in particular it is surrounded by different genes. In this case the nucleic acid molecule may be either under the control of its own promoter or under the control of a heterologous promoter. The vector or nucleic acid molecule according to the invention which is present in the host cell may either be integrated into the genome of the host cell or it may be maintained in some form extrachromosomally. In this respect, it is also to be understood that the nucleic acid molecule of the invention can be used to restore or create a mutant gene via homologous recombination (Paszkowski (ed.), Homologous Recombination and Gene Silencing in Plants. Kluwer Academic Publishers (1994)).

The host cell can be any prokaryotic or eukaryotic cell, such as bacterial, insect, fungal, plant or animal cells. Preferred fungal cells are, for example, those of the genus Saccharomyces, in particular those of the species S. cerevisiae.

The term "prokaryotic" is meant to include all bacteria which can be transformed or transfected with a DNA or RNA molecules for the expression of a protein of the invention. Prokaryotic hosts may include gram negative as well as gram positive bacteria such as, for example, E. coli, S. typhimurium, Serratia marcescens and Bacillus subtilis. The term "eukaryotic" is meant to include yeast, higher plant, insect and preferably mammalian cells. Depending upon the host employed in a recombinant production procedure, the protein encoded by the polynucleotide of the present invention may be glycosylated or may be non-glycosylated. The cell cycle interacting proteins of the invention may or may not also include an initial methionine amino acid residue. A polynucleotide of the invention can be used to transform or transfect the host using any of the techniques commonly known to those of ordinary skill in the art. Furthermore, methods for preparing fused, operably linked genes and expressing them in, e.g., mammalian cells and bacteria are well-known in the art (Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).

Another subject of the invention is a method for the preparation of cell cycle interacting proteins which comprises the cultivation of host cells according to the invention which, due to the presence of a vector or a nucleic acid molecule according to the invention, are able to express such a protein, under conditions which allow expression of the

protein and recovering of the so-produced protein from the culture. It is also to be understood that the proteins can be expressed in a cell free system using for example in vitro translation assays known in the art.

The term "expression" means the production of a protein or nucleotide sequence in the cell. However, said term also includes expression of the protein in a cell-free system. It includes transcription into an RNA product, post-transcriptional modification and/or translation to a protein product or polypeptide from a DNA encoding that product, as well as possible post-translational modifications. Depending on the specific constructs and conditions used, the protein may be recovered from the cells, from the culture medium or from both. The terms "protein" and "polypeptide" used in this application are interchangeable. "Polypeptide" refers to a polymer of amino acids (amino acid sequence) and does not refer to a specific length of the molecule. Thus peptides and oligopeptides are included within the definition of polypeptide. This term does also refer to or include post-translational modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. For example, it is well known by the person skilled in the art that it is not only possible to express a native protein but also to express the protein as fusion polypeptides or to add signal sequences directing the protein to specific compartments of the host cell, e.g., ensuring secretion of the protein into the culture medium, etc. The protein of the invention may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor XA or enterokinase (Invitrogen, San Diego CA) between the purification domain and the protein of interest is useful to facilitate purification. One such expression vector provides for expression of a fusion protein compromising a cell cycle interacting protein and

contains nucleic acid encoding 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography as described in Porath, Protein Expression and Purification 3 (1992), 263-281) while the enterokinase cleavage site provides a means for purifying the cell cycle interacting protein from the fusion protein. In addition to recombinant production, fragments of the protein of the invention may be produced by direct peptide synthesis using solid-phase techniques (cf Stewart et al. (1969) Solid Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield, J. Am. Chem. Soc. 85 (1963), 2149-2154). *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City CA) in accordance with the instructions provided by the manufacturer. Various fragments of the cell cycle interacting protein of the invention may be chemically synthesized and/or modified separately and combined using chemical methods to produce the full length molecule. Once expressed or synthesized, the protein of the present invention can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like; see, Scopes, "Protein Purification", Springer-Verlag, N.Y. (1982). Substantially pure proteins of at least about 90 to 95% homogeneity are preferred, and 98 to 99% or more homogeneity are most preferred, for pharmaceutical uses. Once purified, partially or to homogeneity as desired, the proteins may then be used therapeutically (including extracorporeally) or in developing and performing assay procedures.

Cell cycle interacting proteins of the invention

The present invention furthermore relates to cell cycle interacting proteins encoded by the nucleic acid molecules according to the invention or produced or obtained by the above-described methods, and to functional and/or immunologically active fragments of such cell cycle interacting proteins. The proteins and polypeptides of the present invention are not necessarily translated from a designated nucleic acid sequence; the polypeptides may be generated in any manner, including for example, chemical

synthesis, or expression of a recombinant expression system, or isolation from a suitable viral system. The polypeptides may include one or more analogs of amino acids, phosphorylated amino acids or unnatural amino acids. Methods of inserting analogs of amino acids into a sequence are known in the art. The polypeptides may also include one or more labels, which are known to those skilled in the art. In this context, it is also understood that the proteins according to the invention may be further modified by conventional methods known in the art. By providing the proteins according to the present invention it is also possible to determine fragments which retain biological activity. This allows the construction of chimeric proteins and peptides comprising an amino sequence derived from the protein of the invention, which is crucial for its, e.g., binding activity and other functional amino acid sequences, e.g. GUS marker gene (Jefferson, EMBO J. 6 (1987), 3901-3907). The other functional amino acid sequences may be either physically linked by, e.g., chemical means to the proteins of the invention or may be fused by recombinant DNA techniques well known in the art.

The term "fragment of a sequence" or "part of a sequence" means a truncated sequence of the original sequence referred to. The truncated sequence (nucleic acid or protein sequence) can vary widely in length; the minimum size being a sequence of sufficient size to provide a sequence with at least a comparable function and/or activity of the original sequence referred to, while the maximum size is not critical. In some applications, the maximum size usually is not substantially greater than that required to provide the desired activity and/or function(s) of the original sequence. Typically, the truncated amino acid sequence will range from about 5 to about 60 amino acids in length. More typically, however, the sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

Furthermore, folding simulations and computer redesign of structural motifs of the protein of the invention can be performed using appropriate computer programs (Olszewski, Proteins 25 (1996), 286-299; Hoffman, Comput. Appl. Biosci. 11 (1995), 675-679). Computer modeling of protein folding can be used for the conformational and energetic analysis of detailed peptide and protein models (Monge, J. Mol. Biol. 247 (1995), 995-1012; Renouf, Adv. Exp. Med. Biol. 376 (1995), 37-45). In particular, the appropriate

programs can be used for the identification of interactive sites of the cell cycle interacting protein and its receptor, its ligand or other interacting proteins by computer assistant searches for complementary peptide sequences (Fassina, Immunomethods 5 (1994), 114-120. Further appropriate computer systems for the design of protein and peptides are described in the prior art, for example in Berry, Biochem. Soc. Trans. 22 (1994), 1033-1036; Wodak, Ann. N. Y. Acad. Sci. 501 (1987), 1-13; Pabo, Biochemistry 25 (1986), 5987-5991. The results obtained from the above-described computer analysis can be used for, e.g., the preparation of peptide mimetics of the protein of the invention or fragments thereof. Such pseudopeptide analogues of the natural amino acid sequence of the protein may very efficiently mimic the parent protein (Benkirane, J. Biol. Chem. 271 (1996), 33218-33224). For example, incorporation of easily available achiral (-amino acid residues into a protein of the invention or a fragment thereof results in the substitution of amide bonds by polymethylene units of an aliphatic chain, thereby providing a convenient strategy for constructing a peptide mimetic (Banerjee, Biopolymers 39 (1996), 769-777). Superactive peptidomimetic analogues of small peptide hormones in other systems are described in the prior art (Zhang, Biochem. Biophys. Res. Commun. 224 (1996), 327-331). Appropriate peptide mimetics of the protein of the present invention can also be identified by the synthesis of peptide mimetic combinatorial libraries through successive amide alkylation and testing the resulting compounds, e.g., for their binding and immunological properties. Methods for the generation and use of peptidomimetic combinatorial libraries are described in the prior art, for example in Ostresh, Methods in Enzymology 267 (1996), 220-234 and Dorner, Bioorg. Med. Chem. 4 (1996), 709-715.

Furthermore, a three-dimensional and/or crystallographic structure of the protein of the invention can be used for the design of peptide mimetic inhibitors of the biological activity of the protein of the invention (Rose, Biochemistry 35 (1996), 12933-12944; Rutenber, Bioorg. Med. Chem. 4 (1996), 1545-1558).

Antibodies

Furthermore, the present invention relates to antibodies specifically recognizing a cell cycle interacting protein according to the invention or parts, i.e. specific fragments or epitopes, of such a protein. The antibodies of the invention can be used to identify and isolate other

cell cycle interacting proteins and genes in any organism, preferably plants. These antibodies can be monoclonal antibodies, polyclonal antibodies or synthetic antibodies as well as fragments of antibodies, such as Fab, Fv or scFv fragments etc. Monoclonal antibodies can be prepared, for example, by the techniques as originally described in Köhler and Milstein, Nature 256 (1975), 495, and Galfré, Meth. Enzymol. 73 (1981), 3, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals. Furthermore, antibodies or fragments thereof to the aforementioned peptides can be obtained by using methods which are described, e.g., in Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988; Coligan, "Current Protocols in Immunology", Wiley/Greene, NY (1991). These antibodies can be used, for example, for the immunoprecipitation and immunolocalization of proteins according to the invention as well as for the monitoring of the synthesis of such proteins, for example, in recombinant organisms, and for the identification of compounds interacting with the protein according to the invention. For example, surface plasmon resonance as employed in the BIAcore system can be used to increase the efficiency of phage antibodies selections, yielding a high increment of affinity from a single library of phage antibodies which bind to an epitope of the protein of the invention (Schier, Human Antibodies Hybridomas 7 (1996), 97-105; Malmborg, J. Immunol. Methods 183 (1995), 7-13). In many cases, the binding phenomena of antibodies to antigens is equivalent to other ligand/anti-ligand binding.

Transgenic plants

Plant cell division can conceptually be influenced in four ways: (i) inhibiting or arresting cell division, (ii) maintaining, facilitating or stimulating cell division, (iii) uncoupling DNA synthesis from mitosis and cytokinesis or (iv) uncoupling cell division from intrinsic developmental or external environmental conditions. Modulation of the expression of a cell cycle interacting protein encoded by a nucleotide sequence according to the invention has surprisingly an advantageous influence on plant cell division characteristics, in particular on the disruption of the G1/S and/or G2/M transition and as a result thereof on the total make-up of the plant concerned or parts thereof. An example is that DNA synthesis, or mitosis may be negatively influenced by interfering with the

formation of a cyclin-dependent protein kinase complex. Alternatively, overexpression of the CDK complex interacting protein accelerates reentry into the cell cycle.
The term "cyclin-dependent protein kinase complex" means the complex formed when a, preferably functional, cyclin associates with a, preferably, functional cyclin dependent kinase. Such complexes may be active in phosphorylating proteins and may or may not contain additional protein species.
The term "protein kinase" means an enzyme catalyzing the phosphorylation of proteins.

To analyse the industrial applicabilities of the invention, transformed plants can be made modulating the nucleotide sequence according to the invention. Such an modulation of the new gene(s), proteins or inactivated variants thereof will either positively or negatively have an effect on cell division. Methods to modify the expression levels and/or ratios and/or the activity are known to persons skilled in the art and include for instance overexpression, co-suppression, the use of ribozymes, sense and anti-sense strategies, gene silencing approaches. "Sense strand" refers to the strand of a double-stranded DNA molecule that is homologous to a mRNA transcript thereof. The "anti-sense strand" contains an inverted sequence which is complementary to that of the "sense strand".

Hence, the nucleic acid molecules according to the invention are in particular useful for the genetic manipulation of plant cells in order to modify the characteristics of plants and to obtain plants with modified, preferably with improved or useful phenotypes. Similarly, the invention can also be used to modulate the cell division and the growth of cells, preferentially plant cells, in *in vitro* cultures. Specifically the plant cell division rate and/or the inhibition of a plant cell division can be influenced by overexpression or reducing the expression of a gene encoding a protein according to the invention. Overexpression of a cell cycle interacting protein encoding gene according to the invention promotes cell proliferation, while reducing gene expression arrests cell division or prevents reentry into the cell cycle. Part of the invention is thus the usage of the nucleic acid molecules as mentioned hereinbefore as a negative or positive regulator of cell proliferation. A transformed plant can thus be obtained by transforming a plant cell with a gene encoding a polypeptide concerned or fragment thereof alone or in combination. For this

purpose tissue specific promoters, in one construct or being present as a separate construct in addition to the sequence concerned, can be used. Surprisingly using a polypeptide or fragment thereof according to the invention or using antisense RNA for the gene according to the invention cell division of the meristems of the plant can be manipulated, positively and/or negatively respectively. Furthermore, overproduction of the cell cycle interacting protein of the invention enhances growth and results in cell division to be less sensitive to an arrest caused by environmental stress such as salt, nutrient deprivation, drought, chilling and the like.

Thus, the present invention relates to a method for the production of transgenic plants, plant cells or plant tissue comprising the introduction of a nucleic acid molecule or vector of the invention into the genome of said plant, plant cell or plant tissue.

For the expression of the nucleic acid molecules according to the invention in sense or antisense orientation in plant cells, the molecules are placed under the control of regulatory elements which ensure the expression in plant cells. These regulatory elements may be heterologous or homologous with respect to the nucleic acid molecule to be expressed as well with respect to the plant species to be transformed. In general, such regulatory elements comprise a promoter active in plant cells. These promoters can be used to modulate (e.g. increase, decrease, alter) cell cycle interacting protein content and/or composition in a desired tissue or under certain conditions. To obtain expression in all tissues of a transgenic plant, preferably constitutive promoters are used, such as the 35 S promoter of CaMV (Odell, Nature 313 (1985), 810-812) or promoters from such genes as rice actin (McElroy et al. (1990) Plant Cell 2:163-171) maize H3 histone (Lepetit et al. (1992) Mol. Gen. Genet 231:276-285) or promoters of the polyubiquitin genes of maize (Christensen, Plant Mol. Biol. 18 (1982), 675-689). In order to achieve expression in specific tissues of a transgenic plant it is possible to use tissue specific promoters (see, e.g., Stockhaus, EMBO J. 8 (1989), 2245-2251 or Table A).

Table A: Exemplary tissue specific or tissue-preferred promoters for use in the performance of the present invention.

GENE SOURCE	EXPRESSION PATTERN	REFERENCE
α-amylase (*Amy32b*)	aleurone	Lanahan, M.B., *e t al., Plant Cell 4:*203-211, 1992; Skriver, K., *et al. Proc. Natl. Acad. Sci. (USA) 88:* 7266-7270, 1991
cathepsin β-like gene	aleurone	Cejudo, F.J., *et al. Plant Molecular Biology 20:*849-856, 1992.
Agrobacterium rhizogenes rolB	cambium	Nilsson *et al., Physiol. Plant. 100:*456-462, 1997
PRP genes	cell wall	http://salus.medium.edu/mmg/tierney/html
barley *ltr1* promoter	endosperm	
synthetic promoter	endosperm	Vicente-Carbajosa *et al., Plant J. 13:* 629-640, 1998.
AtPRP4	flowers	http://salus.medium.edu/mmg/tierney/html
chalene synthase (chsA)	flowers	Van der Meer, *et al., Plant Mol. Biol. 15*, 95-109, 1990.
LAT52	anther	Twell et al Mol. Gen Genet. 217:240-245 (1989)
apetala-3	flowers	
chitinase	fruit (berries, grapes, etc)	Thomas *et al.* CSIRO Plant Industry, Urrbrae, South Australia, Australia; http://winetitles.com.au/gwrdc/csh95-1.html
rbcs-3A	green tissue (eg leaf)	Lam, E. *et al., The Plant Cell 2:* 857-866, 1990.; Tucker *et al., Plant Physiol. 113:* 1303-1308, 1992.
leaf-specific genes	leaf	Baszczynski, *et al., Nucl. Acid Res. 16:* 4732, 1988.
AtPRP4	leaf	http://salus.medium.edu/mmg/tierney/html
Pinus cab-6	leaf	Yamamoto *et al., Plant Cell Physiol. 35:*773-778, 1994.
SAM22	senescent leaf	Crowell, *et al., Plant Mol. Biol. 18:* 459-466, 1992.
R. japonicum nif gene	nodule	United States Patent No. 4, 803, 165
B. japonicum nifH gene	nodule	United States Patent No. 5, 008, 194
GmENOD40	nodule	Yang, *et al., The Plant J. 3:* 573-585.
PEP carboxylase (PEPC)	nodule	Pathirana, *et al., Plant Mol. Biol. 20:* 437-450, 1992.
leghaemoglobin (Lb)	nodule	Gordon, *et al., J. Exp. Bot. 44:* 1453-

		1465, 1993.
Tungro bacilliform virus gene	phloem	Bhattacharyya-Pakrasi, *et al, The Plant J. 4:* 71-79, 1992.
sucrose-binding protein gene	plasma membrane	Grimes, *et al., The Plant Cell 4:*1561-1574, 1992.
pollen-specific genes	pollen; microspore	Albani, *et al., Plant Mol. Biol. 15:* 605, 1990; Albani, *et al., Plant Mol. Biol. 16:* 501, 1991)
Zm13	pollen	Guerrero et al Mol. Gen. Genet. 224:161-168 (1993)
apg gene	microspore	Twell et al Sex. Plant Reprod. 6:217-224 (1993)
maize pollen-specific gene	pollen	Hamilton, et al., Plant Mol. Biol. 18: 211-218, 1992.
sunflower pollen-expressed gene	pollen	Baltz, et al., The Plant J. 2: 713-721, 1992.
B. napus pollen-specific gene	pollen;anther; tapetum	Arnoldo, et al., J. Cell. Biochem., Abstract No. Y101, 204, 1992.
root-expressible genes	roots	Tingey, et al., EMBO J. 6: 1, 1987.
tobacco auxin-inducible gene	root tip	Van der Zaal, et al., Plant Mol. Biol. 16, 983, 1991.
β-tubulin	root	Oppenheimer, *et al., Gene 63:* 87, 1988.
tobacco root-specific genes	root	Conkling, et al., Plant Physiol. 93: 1203, 1990.
B. napus G1-3b gene	root	United States Patent No. 5, 401, 836
SbPRP1	roots	Suzuki et al., Plant Mol. Biol. 21: 109-119, 1993.
AtPRP1; AtPRP3	roots; root hairs	http://salus.medium.edu/mmg/tierney/html
RD2 gene	root cortex	http://www2.cnsu.edu/ncsu/research
TobRB7 gene	root vasculature	http://www2.cnsu.edu/ncsu/research
AtPRP4	leaves; flowers; lateral root primordia	http://salus.medium.edu/mmg/tierney/html
seed-specific genes	seed	Simon, *et al., Plant Mol. Biol. 5:* 191, 1985; Scofield, *et al., J. Biol. Chem. 262:* 12202, 1987.; Baszczynski, *et al., Plant Mol. Biol. 14:* 633, 1990.
Brazil Nut albumin	seed	Pearson, et al., Plant Mol. Biol. 18: 235-245, 1992.
legumin	seed	Ellis, et al., Plant Mol. Biol. 10: 203-214,

		1988.
glutelin (rice)	seed	Takaiwa, et al., Mol. Gen. Genet. 208: 15-22, 1986; Takaiwa, et al., FEBS Letts. 221: 43-47, 1987.
zein	seed	Matzke et al Plant Mol Biol, 14(3):323-32 1990
napA	seed	Stalberg, *et al, Planta 199:* 515-519, 1996.
sunflower oleosin	seed (embryo and dry seed)	Cummins, et al., Plant Mol. Biol. 19: 873-876, 1992
LEAFY	shoot meristem	Weigel *et al., Cell 69*:843-859, 1992.
Arabidopsis thaliana knat1	shoot meristem	Accession number AJ131822
Malus domestica kn1	shoot meristem	Accession number Z71981
CLAVATA1	shoot meristem	Accession number AF049870
stigma-specific genes	stigma	Nasrallah, et al., Proc. Natl. Acad. Sci. USA 85: 5551, 1988; Trick, *et al., Plant Mol. Biol. 15:* 203, 1990.
class I patatin gene	tuber	Liu et al., Plant Mol. Biol. 153:386-395, 1991.
blz2	endosperm	EP99106056.7
PCNA rice	meristem	Kosugi *et al, Nucleic Acids Research 19:*1571-1576, 1991; Kosugi S. and Ohashi Y, *Plant Cell 9:*1607-1619, 1997.

The promoters listed in the table are provided for the purposes of exemplification only and the present invention is not to be limited by the list provided therein. Those skilled in the art will readily be in a position to provide additional promoters that are useful in performing the present invention. The promoters listed may also be modified to provide specificity of expression as required.

Known are also promoters which are specifically active in tubers of potatoes or in seeds of different plants species, such as maize, Vicia, wheat, barley etc. Inducible promoters may be used in order to be able to exactly control expression under certain environmental or developmental conditions such as pathogens, anaerobia, light, etc. An example for inducible promoters are the promoters of genes encoding heat shock proteins. Also microspore-specific regulatory elements and their uses have been described

(WO96/16182). Furthermore, the chemically inducible Tet-system may be employed (Gatz, Mol. Gen. Genet. 227 (1991); 229-237). Further suitable promoters are known to the person skilled in the art and are described, e.g., in Ward (Plant Mol. Biol. 22 (1993), 361-366). The regulatory elements may further comprise transcriptional and/or translational enhancers functional in plants cells. Furthermore, the regulatory elements may include transcription termination signals, such as a poly-A signal, which lead to the addition of a poly A tail to the transcript which may improve its stability.

In the case that a nucleic acid molecule according to the invention is expressed in sense orientation it is in principle possible to modify the coding sequence in such a way that the protein is located in any desired compartment of the plant cell. These include the nucleus, endoplasmatic reticulum, the vacuole, the mitochondria, the plastids, the apoplast, the cytoplasm etc. Since the interacting component of the protein of the invention excerts its effects in the cytoplasm and/or nucleus, corresponding signal sequences are preferred to direct the protein of the invention in the same compartment. Methods how to carry out this modifications and signal sequences ensuring localization in a desired compartment are well known to the person skilled in the art.

Methods for the introduction of foreign DNA into plants are also well known in the art. These include, for example, the transformation of plant cells or tissues with T-DNA using Agrobacterium tumefaciens or Agrobacterium rhizogenes, the fusion of protoplasts, direct gene transfer (see, e.g., EP-A 164 575), injection, electroporation, biolistic methods like particle bombardment, pollen-mediated transformation, plant RNA virus-mediated transformation, liposome-mediated transformation, transformation using wounded or enzyme-degraded immature embryos, or wounded or enzyme-degraded embryogenic callus and other methods known in the art. The vectors used in the method of the invention may contain further functional elements, for example "left border"- and "right border"-sequences of the T-DNA of Agrobacterium which allow for stably integration into the plant genome. Furthermore, methods and vectors are known to the person skilled in the art which permit the generation of marker free transgenic plants, i.e. the selectable or scorable marker gene is lost at a certain stage of plant development or plant breeding. This can be achieved by, for example cotransformation (Lyznik, Plant Mol.

Biol. 13 (1989), 151-161; Peng, Plant Mol. Biol. 27 (1995), 91-104) and/or by using systems which utilize enzymes capable of promoting homologous recombination in plants (see, e.g., WO97/08331; Bayley, Plant Mol. Biol. 18 (1992), 353-361); Lloyd, Mol. Gen. Genet. 242 (1994), 653-657; Maeser, Mol. Gen. Genet. 230 (1991), 170-176; Onouchi, Nucl. Acids Res. 19 (1991), 6373-6378). Methods for the preparation of appropriate vectors are described by, e.g., Sambrook (Molecular Cloning; A Laboratory Manual, 2nd Edition (1989), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Suitable strains of Agrobacterium tumefaciens and vectors as well as transformation of Agrobacteria and appropriate growth and selection media are well known to those skilled in the art and are described in the prior art (GV3101 (pMK90RK), Koncz, Mol. Gen. Genet. 204 (1986), 383-396; C58C1 (pGV 3850kan), Deblaere, Nucl. Acid Res. 13 (1985), 4777; Bevan, Nucleic. Acid Res. 12(1984), 8711; Koncz, Proc. Natl. Acad. Sci. USA 86 (1989), 8467-8471; Koncz, Plant Mol. Biol. 20 (1992), 963-976; Koncz, Specialized vectors for gene tagging and expression studies. In: Plant Molecular Biology Manual Vol 2, Gelvin and Schilperoort (Eds.), Dordrecht, The Netherlands: Kluwer Academic Publ. (1994), 1-22; EP-A-120 516; Hoekema: The Binary Plant Vector System, Offsetdrukkerij Kanters B.V., Alblasserdam (1985), Chapter V, Fraley, Crit. Rev. Plant. Sci., 4, 1-46; An, EMBO J. 4 (1985), 277-287). Although the use of Agrobacterium tumefaciens is preferred in the method of the invention, other Agrobacterium strains, such as Agrobacterium rhizogenes, may be used, for example if a phenotype conferred by said strain is desired.

Methods for the transformation using biolistic methods are well known to the person skilled in the art; see, e.g., Wan, Plant Physiol. 104 (1994), 37-48; Vasil, Bio/Technology 11 (1993), 1553-1558 and Christou (1996) Trends in Plant Science 1, 423-431. Microinjection can be performed as described in Potrykus and Spangenberg (eds.), Gene Transfer To Plants. Springer Verlag, Berlin, NY (1995).

The transformation of most dicotyledonous plants is possible with the methods described above. But also for the transformation of monocotyledonous plants several successful transformation techniques have been developed. These include the transformation using biolistic methods as, e.g., described above as well as protoplast transformation, electroporation of partially permeabilized cells, introduction of DNA using glass fibers, etc.

The term "transformation" as used herein, refers to the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for the transfer. The polynucleotide may be transiently or stably introduced into the host cell and may be maintained non-integrated, for example, as a plasmid or as chimeric links, or alternatively, may be integrated into the host genome. The resulting transformed plant cell can then be used to regenerate a transformed plant in a manner known by a skilled person.

In general, the plants which can be modified according to the invention and which either show overexpression of a protein according to the invention or a reduction of the synthesis of such a protein can be derived from any desired plant species. They can be monocotyledonous plants or dicotyledonous plants, preferably they belong to plant species of interest in agriculture, wood culture or horticulture interest, such as crop plants (e.g. maize, rice, barley, wheat, rye, oats etc.), potatoes, oil producing plants (e.g. oilseed rape, sunflower, pea nut, soy bean, etc.), cotton, sugar beet, sugar cane, leguminous plants (e.g. beans, peas etc.), wood producing plants, preferably trees, etc.

Thus, the present invention relates also to a transgenic plant cell which contains (preferably stably integrated into its genome) a nucleic acid molecule according to the invention linked to regulatory elements which allow expression of the nucleic acid molecule in plant cells and wherein the nucleic acid molecule is foreign to the transgenic plant cell. For the meaning of foreign; see supra. The presence and expression of the nucleic acid molecule in the transgenic plant cells leads to the synthesis of a cell cycle interacting protein and leads to physiological and phenotypic changes in plants containing such cells.

Thus, the present invention also relates to transgenic plants and plant tissue comprising transgenic plant cells according to the invention. Due to the (over)expression of a cell cycle interacting protein of the invention, e.g., at developmental stages and/or in plant tissue in which they do not naturally occur these transgenic plants may show various physiological, developmental and/or morphological modifications in comparison to wild-type plants.
Therefore, part of this invention is the use of cell cycle genes and/or cell cycle interacting proteins to modulate the level of cell cycle interacting proteins and/or plant cell division

and/or growth in plant cells, plant tissues, plant organs and/or whole plants. To the scope of the invention also belongs a method to influence the activity of cyclin-dependent protein kinase in a plant cell by transforming the plant cell with a nucleic acid molecule according to the invention and/or manipulation of the expression of said molecule. More in particular using a nucleic acid molecule according to the invention, the disruption of plant cell cycle can be accomplished by interfering in the expression of a substrate for cyclin-dependent protein kinase. The latter goal may be achieved, for example, with methods for reducing the amount of active cell cycle interacting proteins.

For example, to obtain transgenic plants overexpressing a A. thaliana cell cycle interacting gene of the invention, its coding region can be cloned, e.g., into the pAT7002 vector (Aoyama and Chua, Plant J. 11 (1997), 605-612). This vector allows inducible expression of the cloned inserts by the addition of the glucocorticoid dexamethasone. For example, following a polymerase chain reaction (PCR) technology the coding region of the cell cycle interacting gene can be amplified using appropriate primers, whereby a first primer contains an XhoI and a second primer contains an SpeI restriction site. The obtained PCR fragment can be purified and cut with XhoI and SpeI. Subsequently the fragment can be cloned into the XhoI and SpeI sites of pTA7002. The resulted binary vector can be transferred into Agrobacterium tumefaciens. This strain can be used to transform Nicotiana tabacum cv. Petit havana using, e.g., the leaf disk protocol (Horsh, Science 227 (1985), 1229-1231) and Arabidopsis thaliana using, e.g., the root transformation protocol (Valvekens, PNAS 85 (1988), 5536-5540). Transgenic plants can then be selected on hygromycine 20 mg/l. Plants can be tested for the gene of interest inducible expression as follows. 2 to 3 leaves of each transformant can be cut in two. Each half can be either submersed in 50 mM Na-citrate buffer (pH 5.8) with or without dexamethasone (0.03 mM concentration). After 24 hours of induction RNA can be extracted from these leaves using the Trizol reagents (Gibco-BRL) according to the manufactures and a northern gel can be run using, e.g., 5 μg of RNA. The gel can be blotted on a nitro-cellulose filter (HybondN+, Amersham) and hybridised with a gene specific probe. Furthermore, seeds of transformants can be put on ½ MS medium with 1% sucrose, both with and without dexamethasone. As a control SR1 seeds should be included. In the presence of dexamethasone the growth behaviour of the transgenic

plants as compared to the control plants is expected to be modified. For example, these transgenic plants may grow faster and/or have additional cells. Furthermore, said plant may be less sensitive to environmental stress compared to the corresponding wild type plant.

Furthermore, the invention also relates to a transgenic plant cell which contains (preferably stably integrated into its genome) a nucleic acid molecule according to the invention or part thereof, wherein the transcription and/or expression of the nucleic acid molecule or part thereof leads to reduction of the synthesis of a cell cycle interacting protein. In a preferred embodiment, the reduction is achieved by an anti-sense, sense, ribozyme, co-suppression and/or dominant mutant effect.

"Antisense" and "antisense nucleotides" means DNA or RNA constructs which block the expression of the naturally occurring gene product.

The provision of the nucleic acid molecules according to the invention opens up the possibility to produce transgenic plant cells with a reduced level of the protein as described above and, thus, with a defect in the cell cycle. Techniques how to achieve this are well known to the person skilled in the art. These include, for example, the expression of antisense-RNA, ribozymes, of molecules which combine antisense and ribozyme functions and/or of molecules which provide for a co-suppression effect; see also supra. When using the antisense approach for reduction of the amount of cell cycle interacting proteins in plant cells, the nucleic acid molecule encoding the antisense-RNA is preferably of homologous origin with respect to the plant species used for transformation. However, it is also possible to use nucleic acid molecules which display a high degree of homology to endogenously occurring nucleic acid molecules encoding a cell cycle interacting protein. In this case the homology is preferably higher than 80%, particularly higher than 90% and still more preferably higher than 95%. The reduction of the synthesis of a protein according to the invention in the transgenic plant cells can result in an alteration in, e.g., cell division. In transgenic plants comprising such cells this can lead to various physiological, developmental and/or morphological changes, preferably to improved regeneration and transformation capacity of, e.g., cultured cells or wounded tissue.

Thus, the present invention also relates to transgenic plants comprising the above-described transgenic plant cells. These may show, for example, a deficiency in cell division and/or reduced growth characteristics compared to wild type plants due to the stable or transient presence of a foreign DNA resulting in at least one of the following features:

(a) disruption of (an) endogenous gene(s) encoding a protein of the invention;

(b) expression of at least on antisense RNA and/or ribozyme against a transcript comprising a nucleic acid molecule of the invention;

(c) expression of a sense and/or non-translatable mRNA of the nucleic acid molecule of the invention;

(d) expression of an antibody of the invention;

(e) incorporation of a functional or non-functional copy of the regulatory sequence of the invention; or

(f) incorporation of a recombinant DNA molecule or vector of the invention.

The present invention also relates to cultured plant tissues comprising transgenic plant cells as described above which either show overexpression of a protein according to the invention or a reduction in synthesis of such a protein.

Any transformed plant obtained according to the invention can be used in a conventional breeding scheme or in *in vitro* plant propagation to produce more transformed plants with the same characteristics and/or can be used to introduce the same characteristic in other varieties of the same or related species. Such plants are also part of the invention. Seeds obtained from the transformed plants genetically also contain the same characteristic and are part of the invention. As mentioned before, the present invention is in principle applicable to any plant and crop that can be transformed with any of the transformation method known to those skilled in the art and includes for instance corn, wheat, barley, rice, oilseed crops, cotton, tree species, sugar beet, cassava, tomato, potato, numerous other vegetables, fruits.

In yet another aspect, the invention also relates to harvestable parts and to propagation material of the transgenic plants according to the invention which either contain transgenic plant cells expressing a nucleic acid molecule according to the invention or which contain

cells which show a reduced level of the described protein. Harvestable parts can be in principle any useful parts of a plant, for example, flowers, pollen, seedlings, tubers, leaves, stems, fruit, seeds, roots etc. Propagation material includes, for example, seeds, fruits, cuttings, seedlings, tubers, rootstocks etc.

Regulatory sequences of cell cycle interacting genes

As mentioned above, the regulatory sequences that naturally drive the expression of the above described cell cycle interacting proteins may prove useful for the expression of heterologous DNA sequences in certain plant tissues and/or at different developmental stages in plant development.

Accordingly, in a further aspect the present invention relates to a regulatory sequence of a promoter naturally regulating the expression of a nucleic acid molecule of the invention described above or of a nucleic acid molecule homologous to a nucleic acid molecule of the invention. With methods well known in the art it is possible to isolate the regulatory sequences of the promoters that naturally regulate the expression of the above-described DNA sequences; see, e.g., Example 8. For example, using the above described nucleic acid molecules as probes a genomic library consisting of plant genomic DNA cloned into phage or bacterial vectors can be screened by a person skilled in the art. Such a library consists e.g. of genomic DNA prepared from seedlings, fractionized in fragments ranging from 5 kb to 50 kb, cloned into the lambda GEM11 (Promega) phages. Phages hybridizing with the probes can be purified. From the purified phages DNA can be extracted and sequenced. Having isolated the genomic sequences corresponding to the genes encoding the above-described cell cycle interacting proteins, it is possible to fuse heterologous DNA sequences to these promoters or their regulatory sequences via transcriptional or translational fusions well known to the person skilled in the art. In order to identify the regulatory sequences and specific elements of these cell cycle genes, 5'-upstream genomic fragments can be cloned in front of marker genes such as *luc*, *gfp* or the GUS coding region and the resulting chimeric genes can be introduced by means of Agrobacterium tumefaciens mediated gene transfer into plants or transfected into plant cells or plant tissue for

transient expression. The expression pattern observed in the transgenic plants or transfected plant cells containing the marker gene under the control of the regulatory sequences of the invention reveal the boundaries of the promoter and its regulatory sequences. Preferably, said regulatory sequence is capable of conferring expression of a heterologous DNA sequence in main and lateral root meristems, shoot apical meristems, embryos at the globular, heart and torpedo stages, floral meristems and/or cambial cells in the stem.

In context with the present invention, the term "regulatory sequence" refers to sequences which influence the specificity and/or level of expression, for example in the sense that they confer cell and/or tissue specificity; see supra. Such regions can be located upstream of the transcription initiation site, but can also be located downstream of it, e.g., in transcribed but not translated leader sequences.

The term "promoter", within the meaning of the present invention refers to nucleotide sequences necessary for transcription initiation, i.e. RNA polymerase binding, and may also include, for example, the TATA box.

The term "nucleic acid molecule homologous to a nucleic acid molecule of the invention", as used herein includes promoter regions and regulatory sequences of other cell cycle interacting protein encoding genes, such as genes from other species, for example, maize, alfalfa, potato, sorghum, millet, coix, barley, wheat and rice the coding region of which share substantial homology to the cell cycle interacting proteins of the invention and which display substantially the same expression pattern. Such promoters are characterized by their capability of conferring expression of a heterologous DNA sequence in meristematic tissue and cells and other tissues mentioned above.

Thus, according to the present invention, regulatory sequences from any species can be used that are functionally homologous to the regulatory sequences of the promoter of the above defined nucleic acid molecules, or promoters of genes that display an identical or similar pattern of expression, in the sense of being expressed in the above-mentioned tissues and cells. However, the expression conferred by the regulatory sequences of the invention may not be limited to, for example, root meristem cells but can include or be restricted to, for example, subdomains of meristems. The particular expression pattern may also depend on the plant/vector system employed. However,

expression of heterologous DNA sequences driven by the regulatory sequences of the invention predominantly occurs in the meristem unless certain elements of the regulatory sequences of the invention, were taken and designed by the person skilled in the art to control the expression of a heterologous DNA sequence in other cell types.

It is also immediately evident to the person skilled in the art that further regulatory elements may be added to the regulatory sequences of the invention. For example, transcriptional enhancers and/or sequences which allow for induced expression of the regulatory sequences of the invention may be employed. A suitable inducible system is for example tetracycline-regulated gene expression as described, e.g., by Gatz, supra.

The regulatory sequence of the invention may preferably be derived from the above described cell cycle interacting genes. Plants that may be suitable sources for such genes have been described above.

Usually, said regulatory sequence is part of a recombinant DNA molecule. In a preferred embodiment of the present invention, the regulatory sequence in the recombinant DNA molecule is operatively linked to a heterologous DNA sequence.
The term heterologous with respect to the DNA sequence being operatively linked to the regulatory sequence of the invention means that said DNA sequence is not naturally linked to the regulatory sequence of the invention. Expression of said heterologous DNA sequence comprises transcription of the DNA sequence, preferably into a translatable mRNA. Regulatory elements ensuring expression in eukaryotic cells, preferably plant cells, are well known to those skilled in the art. They usually comprise poly-A signals ensuring termination of transcription and stabilization of the transcript, see also supra. Additional regulatory elements may include transcriptional as well as translational enhancers; see supra.

In a preferred embodiment, the heterologous DNA sequence of the above-described recombinant DNA molecules encodes a peptide, protein, antisense RNA, sense RNA and/or ribozyme. The recombinant DNA molecule of the invention can be used alone or as part of a vector to express heterologous DNA sequences, which, e.g., encode

proteins for, e.g., the control of disease resistance, modulation of nutrition value or diagnostics of cell cycle related gene expression. The recombinant DNA molecule or vector containing the DNA sequence encoding a protein of interest is introduced into the cells which in turn produce the RNA and optionally protein of interest. For example, the regulatory sequences of the invention can be operatively linked to a lethal gene for use in the production of male and female sterility in plants. Suitable lethal genes include the Bacillus amyloliquefaciens ribonuclease (Hartlet, J. Mol. Biol. 89 (1985)) and the Bacillus amyloliquefaciens ribonuclease expressed with or without its inhibitor, barstar. Another example for a lethal gene is the catalytic A fragment of diphteria toxin (Tweeten, J. Bacteriol. 156 (1983), 680-685). Expression of diphteria toxin within yeast cells causes ADP-ribosylation of elongation factor 2, which leads to inhibition of protein synthesis and eventual cell death (Mattheakis, Mol. Cell. Biol. 12 (1992), 4026-4037).

On the other hand, said protein can be a scorable marker, e.g., luciferase, green fluorescent protein or ß-galactosidase. This embodiment is particularly useful for simple and rapid screening methods for compounds and substances described herein below capable of modulating cell cycle interacting protein gene expression. For example, a cell suspension can be cultured in the presence and absence of a candidate compound in order to determine whether the compound affects the expression of genes which are under the control of regulatory sequences of the invention, which can be measured, e.g., by monitoring the expression of the above-mentioned marker. It is also immediately evident to those skilled in the art that other marker genes may be employed as well, encoding, for example, a selectable marker which provides for the direct selection of compounds which induce or inhibit the expression of said marker.

The regulatory sequences of the invention may also be used in methods of antisense approaches. The antisense RNA may be a short (generally at least 10, preferably at least 14 nucleotides, and optionally up to 100 or more nucleotides) nucleotide sequence formulated to be complementary to a portion of a specific mRNA sequence and/or DNA sequence of the gene of interest. Standard methods relating to antisense technology have been described; see, e.g., Klann, Plant Physiol. 112 (1996), 1321-1330. Following transcription of the DNA sequence into antisense RNA, the antisense RNA binds to its

target sequence within a cell, thereby inhibiting translation of the mRNA and down-regulating expression of the protein encoded by the mRNA. Thus, in a further embodiment, the invention relates to nucleic acid molecules of at least 15 nucleotides in length hybridizing specifically with a regulatory sequence as described above or with a complementary strand thereof. For the possible applications of such nucleic acid molecules, see supra.

The present invention also relates to vectors, particularly plasmids, cosmids, viruses and bacteriophages used conventionally in genetic engineering that comprise a recombinant DNA molecule of the invention. Preferably, said vector is an expression vector and/or a vector further comprising a selection marker for plants. For example of suitable selector markers, see supra. Methods which are well known to those skilled in the art can be used to construct recombinant vectors; see, for example, the techniques described in Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, the recombinant DNA molecules and vectors of the invention can be reconstituted into liposomes for delivery to target cells; see also supra.

The present invention furthermore relates to host cells transformed with a regulatory sequence, a DNA molecule or vector of the invention. Said host cell may be a prokaryotic or eukaryotic cell; see supra.

In a further preferred embodiment, the present invention provides for a method for the production of transgenic plants, plant cells or plant tissue comprising the introduction of a nucleic acid molecule, recombinant DNA molecule or vector of the invention into the genome of said plant, plant cell or plant tissue. For the expression of the heterologous DNA sequence under the control of the regulatory sequence according to the invention in plant cells, further regulatory sequences such as poly A tail may be fused, preferably 3' to the heterologous DNA sequence, see also supra. Further possibilities might be to add Matrix Attachment Sites at the borders of the transgene to act as "delimiters" and insulate against methylation spread from nearby heterochromatic sequences. Methods

for the introduction of foreign DNA into plants, plant cells and plant tissue are described above.

Thus, the present invention relates also to transgenic plant cells which contain stably integrated into the genome a recombinant DNA molecule or vector according to the invention.

Furthermore, the present invention also relates to transgenic plants and plant tissue comprising the above-described transgenic plant cells. These plants may show, for example, modified architecture, increased yield or an increased tolerance to diseases, e.g., nematodes, geminiviruses or to stresses, e.g., salt, heat, nutrient deprivation, etc.
In yet another aspect the invention also relates to harvestable parts and to propagation material of the transgenic plants according to the invention which contain transgenic plant cells described above. Harvestable parts and propagation material can be in principle any useful part of a plant; see supra.

With the regulatory sequences of the invention, it will be possible to study *in vivo* gene expression related to cell cycle interacting proteins. Furthermore, since cell cycle interacting protein expression has different patterns in different stages of physiological and pathological conditions, it is now possible to determine further regulatory sequences which may be important for the up- or down-regulation of the expression or activity of cell cycle interacting proteins, for example in response to ions or elicitors. In addition, it is now possible to *in vivo* study mutations which affect different functional or regulatory aspects of specific gene expression in the cell cycle. Thus, the present invention also relates to the use of the above described regulatory sequences and recombinant DNA molecules of the invention for the expression of heterologous DNA sequences.

The *in vivo* studies referred to above will be suitable to further broaden the knowledge on the mechanisms and genes involved in the control of the cell cycle. Expression of heterologous genes or antisense RNA under the control of the regulatory sequence of the present invention in plants and plant cells may allow the understanding of the function of each of these genes in the plant.

As mentioned hereinbefore, the nucleic acid molecules and proteins of the present invention provide a basis for the development of mimetic compounds that may be inhibitors or activators of cell cycle interacting proteins or their encoding genes. It will be appreciated that the present invention also provides cell based screening methods that allow a high-throughput-screening (HTS) of compounds that may be candidates for such inhibitors and activators.

Thus, the present invention further relates to a method for the identification of an activator or inhibitor of genes encoding cell cycle interacting proteins comprising the steps of:

(a) culturing a plant cell or tissue or maintaining a plant comprising a recombinant DNA molecule comprising a readout system operatively linked to a regulatory sequence of the invention in the presence of a compound or a sample comprising a plurality of compounds under conditions which permit expression of said readout system;

(b) identifying or verifying a sample and compound, respectively, which leads to suppression or activation and/or enhancement of expression of said readout system in said plant, plant cell, or plant tissue.

The present invention further relates to a method for identifying and obtaining an activator or inhibitor of cell cycle interacting proteins comprising the steps of:

(a) combining a compound to be screened with a reaction mixture containing the cell cycle interacting protein of the invention and a readout system capable of interacting with the cell cycle interacting protein under suitable conditions which permit interaction of the cell cycle interacting protein with said readout system;

(b) identifying or verifying a sample and compound, respectively, which leads to suppression or activation of the readout system.

The term "read out system" in context with the present invention means any substrate that can be monitored, for example due to enzymatically induced changes. It also includes DNA sequences which upon transcription and/or expression in a cell, tissue or

organism provide for a scorable and/or selectable phenotype. Such read out systems are well known to those skilled in the art and comprise, for example, substrates for protein kinases, recombinant DNA molecules and marker genes as described above.
The term "plurality of compounds" in a method of the invention is to be understood as a plurality of substances which may or may not be identical.
Said compound or plurality of compounds may be chemically synthesized or microbiologically produced and/or comprised in, for example, samples, e.g., cell extracts from, e.g., plants, animals or microorganisms. Furthermore, said compound(s) may be known in the art but hitherto not known to be capable of suppressing or activating cell cycle interacting proteins. The reaction mixture may be a cell free extract or may comprise a cell or tissue culture. Suitable set ups for the method of the invention are known to the person skilled in the art and are, for example, generally described in Alberts et al., Molecular Biology of the Cell, third edition (1994), in particular Chapter 17. The plurality of compounds may be, e.g., added to the reaction mixture, culture medium, injected into the cell or sprayed onto the plant. The cell or tissue that may be employed in the method of the invention preferably is a host cell, plant cell or plant tissue of the invention described in the embodiments hereinbefore.

If a sample containing a compound or a plurality of compounds is identified in the method of the invention, then it is either possible to isolate the compound from the original sample identified as containing the compound capable of suppressing or activating cell cycle interacting proteins, or one can further subdivide the original sample, for example, if it consists of a plurality of different compounds, so as to reduce the number of different substances per sample and repeat the method with the subdivisions of the original sample. Depending on the complexity of the samples, the steps described above can be performed several times, preferably until the sample identified according to the method of the invention only comprises a limited number of or only one substance(s). Preferably said sample comprises substances of similar chemical and/or physical properties, and most preferably said substances are identical. Several methods are known to the person skilled in the art for producing and screening large libraries to identify compounds having specific affinity for a target. These methods include the phage-display method in which randomized peptides are displayed from phage and

screened by affinity chromatography to an immobilized receptor; see, e.g., WO 91/17271, WO 92/01047, US-A-5,223,409. In another approach, combinatorial libraries of polymers immobilized on a chip are synthesized using photolithography; see, e.g., US-A-5,143,854, WO 90/15070 and WO 92/10092. The immobilized polymers are contacted with a labeled receptor and scanned for label to identify polymers binding to the receptor. The synthesis and screening of peptide libraries on continuous cellulose membrane supports that can be used for identifying binding ligands of the polypeptide of the invention and thus possible inhibitors and activators is described, for example, in Kramer, Methods Mol. Biol. 87 (1998), 25-39. This method can also be used, for example, for determining the binding sites and the recognition motifs in the polypeptide of the invention. In like manner, the substrate specificity of the DnaK chaperon was determined and the contact sites between human interleukin-6 and its receptor; see Rudiger, EMBO J. 16 (1997), 1501-1507 and Weiergraber, FEBS Lett. 379 (1996), 122-126, respectively. Furthermore, the above-mentioned methods can be used for the construction of binding supertopes derived from the polypeptide of the invention. A similar approach was successfully described for peptide antigens of the anti-p24 (HIV-1) monoclonal antibody; see Kramer, Cell 91 (1997), 799-809. A general route to fingerprint analyses of peptide-antibody interactions using the clustered amino acid peptide library was described in Kramer, Mol. Immunol. 32 (1995), 459-465. In addition, antagonists of the polypeptide of the invention can be derived and identified from monoclonal antibodies that specifically react with the polypeptide of the invention in accordance with the methods as described in Doring, Mol. Immunol. 31 (1994), 1059-1067.

More recently, WO 98/25146 described further methods for screening libraries of complexes for compounds having a desired property, especially, the capacity to agonize, bind to, or antagonize a polypeptide or its cellular receptor. The complexes in such libraries comprise a compound under test, a tag recording at least one step in synthesis of the compound, and a tether susceptible to modification by a reporter molecule. Modification of the tether is used to signify that a complex contains a compound having a desired property. The tag can be decoded to reveal at least one step in the synthesis of such a compound. Other methods for identifying compounds which interact with the proteins according to the invention or nucleic acid molecules encoding such molecules are, for example, the in vitro screening with the phage display

system as well as filter binding assays or "real time" measuring of interaction using, for example, the BIAcore apparatus (Pharmacia).
All these methods can be used in accordance with the present invention to identify activators and antagonists of the polypeptide of the invention.

Various sources for the basic structure of such an activator or inhibitor can be employed and comprise, for example, mimetic analogs of the polypeptide of the invention. Mimetic analogs of the polypeptide of the invention or biologically active fragments thereof can be generated by, for example, substituting the amino acids that are expected to be essential for the biological activity with, e.g., stereoisomers, i.e. D-amino acids; see e.g., Tsukida, J. Med. Chem. 40 (1997), 3534-3541. Furthermore, in case fragments are used for the design of biologically active analogs Pro-mimetic components can be incorporated into a peptide to reestablish at least some of the conformational properties that may have been lost upon removal of part of the original polypeptide; see, e.g., Nachman, Regul. Pept. 57 (1995), 359-370. Furthermore, the polypeptide of the invention can be used to identify synthetic chemical peptide mimetics that bind to or can function as a ligand, substrate, binding partner or the receptor of the polypeptide of the invention as effectively as does the natural polypeptide; see, e.g., Engleman, J. Clin. Invest. 99 (1997), 2284-2292.

The structure-based design and synthesis of low-molecular-weight synthetic molecules that mimic the activity of the native biological polypeptide is further described in, e.g., Dowd, Nature Biotechnol. 16 (1998), 190-195; Kieber-Emmons, Current Opinion Biotechnol. 8 (1997), 435-441; Moore, Proc. West Pharmacol. Soc. 40 (1997), 115-119; Mathews, Proc. West Pharmacol. Soc. 40 (1997), 121-125; Mukhija, European J. Biochem. 254 (1998), 433-438.

It is also well known to the person skilled in the art, that it is possible to design, synthesize and evaluate mimetics of small organic compounds that, for example, can act as a substrate or ligand to the polypeptide of the invention. For example, it has been described that D-glucose mimetics of hapalosin exhibited similar efficiency as hapalosin

in antagonizing multidrug resistance assistance-associated protein in cytotoxicity; see Dinh, J. Med. Chem. 41 (1998), 981-987.

The nucleic acid molecule of the invention can also serve as a target for activators and inhibitors. Activators may comprise, for example, proteins that bind to the mRNA of a gene encoding a polypeptide of the invention, thereby stabilizing the native conformation of the mRNA and facilitating transcription and/or translation, e.g., in like manner as Tat protein acts on HIV-RNA. Furthermore, methods are described in the literature for identifying nucleic acid molecules such as an RNA fragment that mimics the structure of a defined or undefined target RNA molecule to which a compound binds inside of a cell resulting in retardation of cell growth or cell death; see, e.g., WO 98/18947 and references cited therein. These nucleic acid molecules can be used for identifying unknown compounds of pharmaceutical and/or agricultural interest, and for identifying unknown RNA targets for use in treating a disease. These methods and compositions can be used in screening for novel antibiotics, bacteriostatics, or modifications thereof or for identifying compounds useful to alter expression levels of proteins encoded by a nucleic acid molecule. Alternatively, for example, the conformational structure of the RNA fragment which mimics the binding site can be employed in rational drug design to modify known antibiotics to make them bind more avidly to the target. One such methodology is nuclear magnetic resonance (NMR), which is useful to identify drug and RNA conformational structures. Still other methods are, for example, the drug design methods as described in WO 95/35367, US-A-5,322,933, where the crystal structure of the RNA fragment can be deduced and computer programs are utilized to design novel binding compounds which can act as antibiotics.

The compounds which can be tested and identified according to a method of the invention may be expression libraries, e.g., cDNA expression libraries, peptides, proteins, nucleic acids, antibodies, small organic compounds, hormones, peptidomimetics, PNAs or the like (Milner, Nature Medicine 1 (1995), 879-880; Hupp, Cell 83 (1995), 237-245; Gibbs, Cell 79 (1994), 193-198 and references cited supra). Furthermore, genes encoding a putative regulator of cell cycle interacting protein and/or which excert their effects up- or downstream the cell cycle interacting protein of the

invention may be identified using, for example, insertion mutagenesis using, for example, gene targeting vectors known in the art (see, e.g., Hayashi, Science 258 (1992), 1350-1353; Fritze and Walden, Gene activation by T-DNA tagging. In *Methods in Molecular* biology 44 (Gartland, K.M.A. and Davey, M.R., eds). Totowa: Human Press (1995), 281-294) or transposon tagging (Chandlee, Physiologia Plantarum 78 (1990), 105-115). Said compounds can also be functional derivatives or analogues of known inhibitors or activators. Such useful compounds can be for example transacting factors which bind to the cell cycle interacting protein or regulatory sequences of the invention. Identification of transacting factors can be carried out using standard methods in the art (see, e.g., Sambrook, supra, and Ausubel, supra). To determine whether a protein binds to the protein or regulatory sequence of the invention, standard native gel-shift analyses can be carried out. In order to identify a transacting factor which binds to the protein or regulatory sequence of the invention, the protein or regulatory sequence of the invention can be used as an affinity reagent in standard protein purification methods, or as a probe for screening an expression library. The identification of nucleic acid molecules which encode proteins which interact with the cell cycle interacting proteins described above can also be achieved, for example, as described in Scofield (Science 274 (1996), 2063-2065) by use of the so-called yeast "two-hybrid system"; see also the appended examples. In this system the protein encoded by the nucleic acid molecules according to the invention or a smaller part thereof is linked to the DNA-binding domain of the GAL4 transcription factor. A yeast strain expressing this fusion protein and comprising a lacZ reporter gene driven by an appropriate promoter, which is recognized by the GAL4 transcription factor, is transformed with a library of cDNAs which will express plant proteins or peptides thereof fused to an activation domain. Thus, if a peptide encoded by one of the cDNAs is able to interact with the fusion peptide comprising a peptide of a protein of the invention, the complex is able to direct expression of the reporter gene. In this way the nucleic acid molecules according to the invention and the encoded peptide can be used to identify peptides and proteins interacting with cell cycle interacting proteins. It is apparent to the person skilled in the art that this and similar systems may then further be exploited for the identification of inhibitors of the binding of the interacting proteins.

Once the transacting factor is identified, modulation of its binding to or regulation of expression of the cell cycle interacting protein of the invention can be pursued,

beginning with, for example, screening for inhibitors against the binding of the transacting factor to the protein of the present invention. Activation or repression of cell cycle interacting proteins could then be achieved in plants by applying of the transacting factor (or its inhibitor) or the gene encoding it, e.g. in a vector for transgenic plants. In addition, if the active form of the transacting factor is a dimer, dominant-negative mutants of the transacting factor could be made in order to inhibit its activity. Furthermore, upon identification of the transacting factor, further components in the pathway leading to activation (e.g. signal transduction) or repression of a gene involved in the control of cell cycle then can be identified. Modulation of the activities of these components can then be pursued, in order to develop additional drugs and methods for modulating the cell cycle in animals and plants.

Thus, the present invention also relates to the use of the two-hybrid system as defined above for the identification of cell cycle interacting proteins or activators or inhibitors of such poteins

Determining whether a compound is capable of suppressing or activating cell cycle interacting proteins can be done, for example, by monitoring DNA duplication and cell division. It can further be done by monitoring the phenotypic characteristics of the cell of the invention contacted with the compounds and compare it to that of wild-type plants. In an additional embodiment, said characteristics may be compared to that of a cell contacted with a compound which is either known to be capable or incapable of suppressing or activating cell cycle interacting proteins.

The compounds isolated by the above methods also serve as lead compounds for the development of analog compounds. The analogs should have a stabilized electronic configuration and molecular conformation that allows key functional groups to be presented to the receptor in substantially the same way as the lead compound. In particular, the analog compounds have spatial electronic properties which are comparable to the binding region, but can be smaller molecules than the lead compound, frequently having a molecular weight below about 2 kD and preferably below about 1 kD. Identification of analog compounds can be performed through use of techniques such as self-consistent field (SCF) analysis, configuration interaction (CI)

analysis, and normal mode dynamics analysis. Computer programs for implementing these techniques are available; e.g., Rein, Computer-Assisted Modeling of Receptor-Ligand Interactions (Alan Liss, New York, 1989). Methods for the preparation of chemical derivatives and analogues are well known to those skilled in the art and are described in, for example, Beilstein, Handbook of Organic Chemistry, Springer edition New York Inc., 175 Fifth Avenue, New York, N.Y. 10010 U.S.A. and Organic Synthesis, Wiley, New York, USA. Furthermore, said derivatives and analogues can be tested for their effects according to methods known in the art. Furthermore, peptidomimetics and/or computer aided design of appropriate derivatives and analogues can be used, for example, according to the methods described above.

The inhibitor or activator identified by the above-described method may prove useful as a herbicide, pesticide, insecticide, antibiotic, tumor suppressing agent and/or as a cell growth regulator. Thus, in a further embodiment the invention relates to a compound obtained or identified according to the method of the invention said compound being an activator of cell cycle interacting proteins or an inhibitor of cell cycle interacting proteins. The above-described compounds include, for example, cell cycle kinase inhibitors. "Cell-cycle kinase inhibitor" (CKI) is a protein which inhibit CDK/cyclin activity and is produced and/or activated when further cell division has to be temporarily or continuously prevented. The antibodies, nucleic acid molecules, inhibitors and activators of the present invention preferably have a specificity at least substantially identical to the binding specificity of the natural ligand or binding partner of the cell cycle protein of the invention, in particular if cell cycle stimulation is desired. An antibody or inhibitor can have a binding affinity to the cell cycle interacting protein of the invention of at least 10^5M^{-1}, preferably higher than 10^7M^{-1} and advantageously up to $10^{10}M^{-1}$ in case cell cycle suppression should be mediated.

In a preferred embodiment, a suppressive antibody or inhibitor of the invention has an affinity of at least about 10^{-7} M, preferably at least about 10^{-9} M and most preferably at least about 10^{-11} M; and cell cycle stimulating activator has an affinity of less than about 10^{-7} M, preferably less than about 10^{-6} M and most preferably in order of 10^{-5}M.

In case of nucleic acid molecules it is preferred that they have a binding affinity to those encoding the amino acid sequences depicted in SEQ ID NO: 2, 4, 34, 36, 38, 40, 42, 6,

8, 10, 12 or 14 of at most 2-, 5- or 10-fold less than an exact complement of 20 consecutive nucleotides of the above described nucleic acid molecules.

Preferably, the compound identified according to the above described method or its analog or derivative is further formulated in a therapeutically active form or in a form suitable for the application in plant breeding or plant cell and tissue culture. For example, it can be combined with a agriculturally acceptable carrier known in the art. Thus, the present invention also relates to a method of producing a therapeutic or plant effective composition comprising the steps of one of the above described methods of the invention and combining the compound obtained or identified in the method of the invention or an analog or derivative thereof with a pharmaceutically acceptable carrier or with a plant cell and tissue culture acceptable carrier. As is evident from the above, the present invention generally relates to compositions comprising at least one of the aforementioned nucleic acid molecules, vectors, proteins, regulatory sequences, recombinant DNA molecules, antibodies or compounds. Advantageously, said composition is for use as a medicament, a diagnostic means, a kit or as a plant effective composition.

Compositions useful in agriculture and in plant cell and tissue culture

Plant protection compositions can be prepared by employing the above-described methods of the invention and synthesizing the compound identified as inhibitor or activator in an amount sufficient for use in agriculture. Thus, the present invention also relates to a method for the preparation of an agricultural plant protection composition comprising the above-described steps of the method of the invention and synthesizing the compound so identified or an analog or derivative thereof.

In the plant protection composition of the invention, the compound identified by the above-described method may be preferentially formulated by conventional means commonly used for the application of, for example, herbicides and pesticides or agents capable of inducing systemic acquired resistance (SAR). For example, certain additives known to those skilled in the art stabilizers or substances which facilitate the uptake by the plant cell, plant tissue or plant may be used.

Pharmaceutical compositions

The cell cycle interacting proteins of the invention appear to function in the cell division cycle which is similar in plants and animals. Accordingly, the nucleic acid molecules and proteins of the invention or derivatives thereof as well as the above described activators and inhibitors may be used to modulate the cell division cycle in animal, preferably mammalian cells which is integral to the development and spread of cancerous cells. A cell cycle interacting protein that acts as a basal transcription factor may promote cancer cell growth. In conditions where cell cycle interacting protein activity is not desirable, cells could be transfected with antisense sequences to cell cycle interacting protein encoding polynucleotides or provided with antagonists to the protein or its encoding gene. Thus, the above described antagonists or antisense molecules may be used to slow, stop, or reverse cancer cell growth. Thus, the present invention also relates to a method of producing a therapeutic agent comprising the steps of the methods described hereinbefore and synthesizing the activator or inhibitor obtained or identified in step (c) or an analog or derivative thereof in an amount sufficient to provide said agent in a therapeutically effective amount to a patient.

Compounds identified by the above methods or analogs are formulated for therapeutic use as pharmaceutical compositions. The compositions can also include, depending on the formulation desired, pharmaceutically acceptable, usually sterile, non-toxic carriers or diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are distilled water, physiological saline, Ringer's solutions, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation may also include other carriers, adjuvants, or nontoxic, nontherapeutic, nonimmunogenic stabilizers and the like. A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors which ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population).

The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50.

Diagnostic means and kits

The invention also relates to a diagnostic composition comprising at least one of the aforementioned nucleic acid molecules, vectors, proteins, antibodies or compounds and optionally suitable means for detection. Said diagnostic compositions may be used for methods for determining expression of cell cycle interacting proteins by detecting the presence of the corresponding mRNA which comprises isolation of mRNA from a cell and contacting the mRNA so obtained with a probe comprising a nucleic acid probe as described above under hybridizing conditions, detecting the presence of mRNA hybridized to the probe, and thereby detecting the expression of the protein in the cell. Further methods of detecting the presence of a protein according to the present invention comprises immunotechniques well known in the art, for example enzyme linked immunosorbent assay. Furthermore, it is possible to use the nucleic acid molecules according to the invention as molecular markers in plant breeding. Moreover, the present invention relates to a kit comprising at least one of the aforementioned nucleic acid molecules, regulatory sequences, recombinant DNA molecules, vectors, proteins, compounds or antibodies of the invention. The kit of the invention may contain further ingredients such as selection markers and components for selective media suitable for the generation of transformed host cells and transgenic plant cells, plant tissue or plants. Furthermore, the kit may include buffers and substrates for reporter genes that may be present in the recombinant gene or vector of the invention. The kit of the invention may advantageously be used for carrying out the method of the invention and could be, inter alia, employed in a variety of applications referred to herein, e.g., in the diagnostic field or as research tool. The parts of the kit of the invention can be packaged individually in vials or in combination in containers or multicontainer units. Manufacture of the kit follows preferably standard procedures which are known to the person skilled in the art. The kit or its ingredients according to the invention can be used in plant cell and plant tissue cultures, for example, for any of the above described methods for detecting inhibitors and activators of cell cycle genes. The kit of the

invention and its ingredients are expected to be very useful in breeding new varieties of, for example, plants which display improved properties such as nutritial value or disease resistance.

Further applications of the invention

The person skilled in the art can use proteins according to the invention from other organisms such as yeast and animals to influence cell division progression in those other organisms such as mammals or insects. In a preferred embodiment one or more DNA sequences, vectors or proteins of the invention or the above-described antibody or compound are, for instance, used to specifically interfere in the modulation of the protein levels or activity of any protein involved in disruption of the expression levels of genes involved in G1/S and/or G2/M transition in the cell cycle process in transformed plants, particularly :

- in the complete plant
- in selected plant organs, tissues or cell types
- under specific environmental conditions, including abiotic stress such as cold, nutrient deprivation, heat, drought or salt stress or biotic stress such as pathogen attack
- during specific developmental stages.

Specifically the plant cell division rate and/or the inhibition of a plant cell division can be influenced by (partial) elimination of a gene or reducing the expression of a gene encoding a protein according to the invention. Said plant cell division rate and/or the inhibition of a plant cell division can also be influenced by eliminating or inhibiting the activity of the protein according to the invention by using for instance antibodies directed against said protein. As a result of said elimination or reduction greater organisms or specific organs or tissues can be obtained; greater in volume and in mass too. Furthermore inhibition of cell division by various adverse environmental conditions such as drought, nutrient deprivation, high salt content, chilling and the like can be delayed or prevented by reduction or enhancing (e.g. with a dominant negative version) of said expression of a gene according to the invention. The division rate of a plant cell can also

be influenced in a transformed plant by overexpression of a nucleic acid molecule according to the invention. Therefore an important aspect of the current invention is a method to modify plant architecture by overproduction or reduction of expression of a sequence according to the invention under the control of a tissue, cell or organ specific promoter. Another aspect of the present invention is a method to modify the growth inhibition of plants caused by environmental stress conditions above mentioned, or more particularly salt stress or nutrient deprivation by appropriate use of sequences according to the invention. Surprisingly using a polypeptide or fragment thereof according to the invention or using antisense RNA or any method to reduce the expression of the gene according to the invention, cell division in the meristem of both main and lateral roots, shoot apical or the vascular tissue of a plant can be manipulated. Furthermore any of the DNA sequences of the invention can be used to manipulate (reduce or enhance) the level of endopolyploidy and thereby increasing the storage capacity of, for example, endosperm cells. Thus, another aspect of the current invention is that one or more DNA sequences, vectors or proteins, regulatory sequences or recombinant DNA molecules of the invention or the above-described antibody or compound can be used to modulate, for instance, endoreduplication in storage cells, storage tissues and/or storage organs of plants or parts thereof. The term "endoreduplication" means recurrent DNA replication without consequent mitosis and cytokinesis.

Preferred target storage organs and parts thereof for the modulation of endoreduplication are, for instance, seeds (such as from cereals, oilseed crops), roots (such as in sugar beet), tubers (such as in potato) and fruits (such as in vegetables and fruit species). Furthermore it is expected that increased endoreduplication in storage organs and parts thereof correlates with enhanced storage capacity and as such with improved yield. In yet another embodiment of the invention, a plant with modulated endoreduplication in the whole plant or parts thereof can be obtained from a single plant cell by transforming the cell, in a manner known to the skilled person, with the above-described means.

In view of the foregoing, the present invention also relates to the use of a DNA sequence, vector, protein, antibody, regulatory sequences, recombinant DNA molecule, nucleic acid molecules or compound of the invention for modulating plant cell cycle, plant cell division and/or growth, for influencing the activity of cell cycle interacting

protein, for disrupting plant cell division by influencing the presence or absence or by interfering in the expression of a cyclin-dependent protein, for modifying growth inhibition of plants caused by environmental stress conditions, for inducing male or female sterility, for influencing cell division progression in a host as defined above or for use in a screening method for the identification of inhibitors or activators of cell cycle proteins.

Furthermore, it is possible to use the nucleic acid molecules according to the invention as molecular markers in plant breeding. Thus, the present invention also relates to the use of a DNA sequence or regulatory sequence of the invention as a marker gene in plant or animal cell and tissue culture or as a marker in marker-assisted plant breeding. Moreover, the overexpression of nucleic acid molecules according to the invention may be useful for the alteration or modification of plant/pathogen interaction. The term "pathogen" includes, for example, bacteria, viruses and fungi as well as protozoa.

Regulation of phosphate assimilation

In a preferred embodiment of the invention, DNA sequences of section (b) encoding PLP proteins described herein before and corresponding vectors, proteins etc. of the invention, in particular PHO80-like proteins (PLPs) may be used to improve the tolerance of plants towards suboptimal nutrient conditions, in particular levels of phosphate. Therefore such sequences may be used to uncouple optimal phosphate conditions from the plant growth rate resulting in enhanced growth rates in normal conditions or stress conditions such as low phosphate. Plants with modified expression of the PLP genes can display enhanced growth rates in normal growth conditions and in different stress conditions, in particular in the case of nutritional deprivation. Plants with modified expression of the PLP genes encompasses a method for conferring plant tolerance towards low levels of phosphate, meaning they may also be useful as a transgenic selective markers.

The cDNA clone (LDV24) was isolated according to the invention as a novel protein interacting with the CDC2aAt protein in a two-hybrid screen. This clone encodes a protein showing strongest homology to the cyclins PHO80, PCL1 and PCL2 from *Saccharomyces cerevisiae* and to the PREG protein from *Neurospora crassa*, and was

renamed PLP5 (PHO80-like protein). Also four cDNAs, named PLP1 to PLP4, were isolated by RT-PCR technology from *Arabidopsis thaliana* according to the invention. Tissue specific expression analysis was performed. Two-hybrid analysis demonstrated all plant PLPs interact with the *A. thaliana* CDKs. Overexpression and antisense constructs were designed and introduced into plants.

Physiological response to phosphate stress

Phosphorus availability is considered one of the major growth-limiting factors for plants in many natural ecosystems. The primary source of phosphorus in soils is inorganic phosphate (Pi). Phosphorous is one of the most important nutrients for all organisms as it is part of many key biomolecules, like DNA, RNA, and lipids. In addition Pi plays an essential role in the energy transfer chain and multiple metabolic pathways (Robinson (1996) *Annals of botany* 77, 179-185). For these reasons, plants have developed several adaptive mechanisms to overcome Pi stress, which involve both morphological and metabolic changes. The most common adaptation under limiting Pi are: (a) morphological adaptations such as root growth and architecture changes or (b) metabolic adaptations are represented by: (i) changes in the respiration rate and phospholipid content of chloroplasts. Phosphate availability affects the thylakoid lipid composition, the relative amount of sulfolipids, and a concomitant decrease in phosphatidylglycerol. Also several enzymes of the glycolytic pathway are altered (Theodorou and Plaxton (1993) *Plant Physiol.* 101, 339-344), (ii) secretion of protons and organic acids. Expression of low-Mr organic acids help to mobilise stores of Pi that are present in the soil as insoluble salts (Nagano and Ashihara (1993) *Plant cell Physiol.* 34, 1219-1228), (iii) synthesis of proteins that include high-affinity Pi transporters, RNases (Bariola et al. (1994) *Plant J* . 6, 673-85) and phophatases (Del Pozo et al. (1999) *Plant J.* 19, 579-589). Phosphate uptake by roots and distribution within the plant are presumed to occur via a phosphate/proton (H^+/orthophosphate) cotransport. The uptake rate is enhanced severalfold by Pi deficiency, and there is evidence Pi deficiency induces a high-affinity Pi transporter in root and leaf cells. (Muchhal and Raghothama (1999) *Proc. Natl. Acad.* 96, 5868-72; Liu et al. (1998) *Plant Physiol* 1998 Jan;116, 91-9). When phosphate is still available in the cell, but not outside, the synthesis of extra- and intracellular (cytoplasmatic and vacuolar) RNases by

Pi starvation is induced (Kock et al. (1998) *Plant Mol Biol* 27, 477-85). Also the stimulation of phosphatase activities in response to Pi starvation are well documented. Both RNases and phosphatases are thought to be involved in both Pi acquisition and recycling, depending on their cellular and subcellular location. In addition, a growing number of phosphorus stress response genes are being identified, but for must of them a clear function has not been found yet.

In plants it appears that the regulation of Pi uptake and transport is likely to operate at a number of different sites, including: (a) uptake of Pi from the external medium by root hairs and epidermal cells, (b) movement of Pi through the cortical cells, (c) loading of Pi into xylem vessels in the root, (d) unloading from the xylem into the shoot cells. At the cellular level, the cytoplasmatic Pi concentration is kept constant while the vacuole stores are more labile (Burleigh and Harrison (1999) *Plant Phys.* 119, 241-248).
After the Pi stress signal is done, the rest of the plant exhibits significant metabolic alteration such as: (a) activation of Pi recycling, (b) alteration of plant respiration rate (alternative pathway of glycolisis and mitochondria electron transport), (c) modification in the photosynthesis and photosynthate partitioning in leaves, (d) changes in Pi flow in the vascular system.

Phosphate signal transduction

The mechanisms that control the acclimation of *Escherichia coli* and *Saccharomyces cerevisiae* to Pi limitation have been extensively studied. In *E.coli* a two component regulatory system governs the transcription of many genes that are responsive to the Pi levels of the environment. In *S. cerevisiae* many mutants (pho series mutants) have been isolated. In this system, transcription of the *PHO5* gene, encoding a repressible acid phosphatase (rAPase), is under the control of the phosphate availability in the medium via a complex network of intracellular regulatory factors that comprises at least five genes: *PHO2, PHO4, PH80, PHO81* and *PHO85*. *PHO2* and *PHO4* encode the activators necessary for transcription of PHO5. When the levels of Pi are high, the PHO4 protein is hyperphosphorylated, impeding its nuclear import (and then the interaction with the PHO2 transcription factor). This phosphorylation is mediated by the PHO80/PHO85 cyclin/CDK complex, thus being negative regulatory factors for the

PHO5 expression. The *PHO85* encodes a non-essential protein kinase with 50% identity to the *CDC28*, and *PHO80* encodes a protein with homology to other yeast cyclins. Unlike the well-understood PHO regulation system in *S.cerevisiae*, the basis on which the plants are able to respond to external phosphate concentration are not yet understood.

Link between cell division control and phosphate nutrition

PHO85 and PHO80 (and related sequences of the PHO80 like *PCL1* and *PCL2*) might have substrates that mediate other responses than phosphate starvation, such as regulation of growth and cell division. This is supported by the observation that the *S. cerevisiae* PHO85 protein can interact with the G1 specific cyclins PCL1 and PCL2 (close homologues to the PHO80). In a yeast strain deficient for the G1 cyclins CLN1 and CLN2, PHO85 is required for G1 progression. This result suggests that PHO85 is involved in a regulatory pathway that links the nutrient status of the cell with cell division activity (Gilliquet and Berben (1993) *FEMS Microbiol Lett* ,108, 333-9).

In plants the relationship between growth, cell division and Pi availability is demonstrated by the observed increase of lateral roots when the Pi concentration in the soil decreases, suggesting low Pi levels act as a mitogenic factor. In tobacco BY-2 cells, cell division is inhibited when the Pi is absent in the medium. Cells deprived of phosphate for 3 days induced cells to semi-synchronously re-enter the cell cycle from a static state (Sano et al. (1999) *Plant Cell Physiol.* 40, 1-8). Phosphate as a limiting factor for the cell division of tobacco BY-2 cells. *Plant Cell Physiol* , 40, 1-8). Both events suggests a clear link between cell cycle regulation and the available nutrient levels. This has more recently been demonstrated using a transgenic approach. Plants overexpressing a membrane associated phosphate apyrase show that an increase in phosphate transport correlates with an enhanced growth phenotype (Thomas et al. (1999) *Plant Phys.* 119, 543-551).

In accordance with the present invention, the plant homolog to the cyclin PHO80 has been isolated for the time. They have also identified a family of such PHO80-like proteins (PLPs). The invention therefore encompasses such nucleotide sequences, proteins and their derivatives, variants and homologs. It also provides transgenic plants comprising

PLPs. Thus, the above described embodiments of the present invention may be preferably performed with PLP nucleic acids and protein, for example as illustrated below.

An embodiment of the invention includes a method for modulating (i.e. increasing or decreasing), in a transgenic plant, the expression of PLP genes. The method comprises transforming a plant cell (as described previously) with a vector comprising a nucleotide sequence of a PLP of the invention. In some embodiments modulating the PLP protein may be by use of a promoter to up or down regulate gene expression or to regulate expression in certain tissues or under certain environmental conditions. In a preferred embodiment a constitutive or root-specific promoter is used.

One embodiment of the invention includes a method for improving the tolerance of plants towards suboptimal nutrient conditions, in particular levels of phosphate, by modulating PLP expression and/or activity. Another embodiment includes a method for improving the growth of plants in normal conditions or suboptimal nutrient conditions, in particular levels of phosphate, by modulating PLP expression and/or activity.

An embodiment of the invention includes a method for providing enhanced rate or frequency of seed germination comprising modulating PLP expression and/or activity.

Also in some embodiments coding regions of the PLP genes can be altered by insertion, deletion, substitution or addition to decrease the activity of the encoded protein.

An embodiment of the invention includes using a PLP gene in combination with one or more another PLP genes. Similarly they may be used in combination with other transgenes that confer another phenotype to the plant. Likewise, it is possible to first confer, improved phosphate sensitivity to a plant in accordance with the method of the invention and to then in an additional step transform such plant in accordance thereof with a further nucleic acid molecule, the presence of which results in another new phenotype characteristic of said plant. Irrespective of the actual performance of transformation, the result of the present invention displays at least two new properties compared to a naturally occurring wild-type plant, that is improved phosphate sensitivity and: a phenotype that is due to the presence

of a further nucleic acid molecule in said plants e.g. herbicide or insectide tolerance, resistance to pathogens, improvement of starch composition and/or production etc.; see also supra.

An embodiment of the invention includes a method for using of PLPs as a positive or negative selectable marker during transformation procedures (Wickert et al. (1998) J. Bacteriology 180 (7):1887-1894). Overexpression of the PLP would mean that it could be used as a positive selectable marker during transformation procedures while antisense/cosuppression means it could be used as a negative selective marker. The selective agent is an antibiotic, preferably hygromycin.

These and other embodiments are disclosed and encompassed by the description and examples of the present invention. Further literature concerning any one of the methods, uses and compounds to be employed in accordance with the present invention may be retrieved from public libraries, using for example electronic devices. For example the public database "Medline" may be utilized which is available on the Internet, for example under http://www.ncbi.nlm.nih.gov/PubMed/medline.html. Further databases and addresses, such as http://www.ncbi.nlm.nih.gov/, http://www.infobiogen.fr/, http://www.fmi.ch/biology/research_tools.html, http://www.tigr.org/, are known to the person skilled in the art and can also be obtained using, e.g., http://www.lycos.com. An overview of patent information in biotechnology and a survey of relevant sources of patent information useful for retrospective searching and for current awareness is given in Berks, TIBTECH 12 (1994), 352-364.

The Figures show:

Figure 1: Expression of the PLP genes in Arabidopsis tissues. A gel blot of RT-PCR from the Arabidopsis tissues indicated and from suspension cultured cell is shown. Total RNA was prepared from these tissues, which were harvested complete from 4 weeks old plants.

The present invention is further illustrated by reference to the following non-limiting examples.

Unless stated otherwise in the examples, all recombinant DNA techniques are performed according to protocols as described in Sambrook et al.. (1989), Molecular Cloning : A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY or in Volumes 1 and 2 of Ausubel et al.. (1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfase (1993) by R.D.D. Croy, jointly published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications (UK).

Example 1: Identification of cell cycle interacting proteins using the two hybrid system with CDC2b as a bait

A two-hybrid screening was performed using as bait a fusion between the GAL4 DNA-binding domain and *CDC2bAt*. Vectors and strains used were provided with the Matchmaker Two-Hybrid System (Clontech, Palo Alto, CA). The bait was constructed by inserting the *CDC2bAt* PCR fragment into the pGBT9 vector. The PCR fragment was created from the cDNA using primers to incorporate *Eco*RI restriction enzyme sites (5'-CGGATCCGAATTCATGGAGAACGAG-3' (SEQ ID NO: 15) and 5'-CGGATCCGAATTCTCAGAACTGAGA-3') (SEQ ID NO: 16). The PCR fragment was cut with *Eco*RI and cloned into the *Eco*RI site of pGBT9, resulting in the plasmid pGBTCDC2B. The GAL4 activation domain cDNA fusion library was obtained from Clontech from mRNA of Arabidopsis thaliana cell suspensions harvested at various growing stages: early exponential, exponential, early stationary, and stationary phase.
For the screening a 1-liter culture of the *Saccharomyces cerevisiae* strain HF7c (*MATa ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3,112 gal4-542 gal80-538 LYS2::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-HIS3 URA3::GAL4*$_{17mers(3x)}$*-CyC1*$_{TATA}$*-LacZ*) was cotransformed with 590 μg pGBTCDC2B, 1100 μg DNA of the library, and 40 mg salmon sperm carrier DNA using the lithium acetate method (Gietz *et al.*., 1992). To estimate the number of independent cotransformants, 1/1000 of the transformation mix

was plated on Leu- and Trp- medium. The rest of the transformation mix was plated on medium to select for histidine prototrophy (Trp-, Leu-, His-). After 5 days of growth at 30°C, the colonies larger than 2 mm were streaked on histidine-lacking medium. A total of 10^7 independent cotransformants were screened for there ability to grow on histidine free medium. A 5-day incubation at 30°C yielded 352 colonies. Of the His^+ colonies the activation domain plasmids were isolated as described (Hoffman and Winston, 1987, Gene 57, 267-272). The hybriZAP™ inserts were PCR amplified using the primers 5'-AGGGATGTTTAATACCACTAC-3' (SEQ ID NO: 17) and 5'-GCACAGTTGAAGTGAACTTGC-3' (SEQ ID NO: 18). PCR fragments were digested with *Alu*I and fractionized on a 2% agarose gel. Plasmid DNA of which the inserts gave rise to different restriction patterns were electroporated into Escherichia coli XL1-Blue, and the DNA sequence of the inserts was determined. Extracted DNA was also used to retransform HF7c to test the specificity of the interaction.

Example 2: Identification of cell cycle interacting proteins using the two hybrid system with CDC2a as a bait

For the identification of cell cycle interacting proteins also a two hybrid system based on GAL4 recognition sites to regulate the expression of both *his3* and *lacZ* reporter genes was used to identify CDC2aAt-interacting of proteins. The bait used for the two-hybrid screening was constructed by inserting the *CDC2aAt* coding region into the pGBT9 vector (Clontech). The insert was created by PCR using the *CDC2aAt* cDNA as template. Primers were designed to incorporate *Eco*RI restriction enzyme sites. The primers used were 5'-CGAGATCTGAATTCATGGATCAGTA-3' (SEQ ID NO: 19) and 5'-CGAGATCTGAATTCCTAAGGCATGCC-3' (SEQ ID NO: 20). The PCR fragment was cut with *Eco*RI and cloned into the *Eco*RI site of pGBT9, resulting in the pGBTCDC2A plasmid. For the screening a GAL4 activation domain cDNA fusion library was used constructed from *Arabidopsis thaliana* cell suspension cultures. This library was constructed using RNA isolated from cells harvested at 20 hours, 3, 7 and 10 days after dilution of the culture in new medium. These time point correspondent to cells from the early exponential growth phase to the late stationary phase. mRNA was prepared using Dynabeads oligo$(dT)_{25}$ according to the manufacturer's instructions (Dynal). The GAL4

activation domain cDNA fusion library was generated using the HybriZAP™ vector purchased with the HybriZAP™ Two-Hybird cDNA Gigapack cloning Kit (Stratagene) following the manufacturer's instructions. The resulting library contained approximately 3.10^6 independent plaque-forming units, with an average insert size of 1 Kb.

For the screening a 1-liter culture of the *Saccharomyces cerevisiae* strain HF7c (*MAT_a ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3,112 gal4-542 gal80-538 LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3 URA3::GAL4$_{17mers(3x)}$-CyC1$_{TATA}$-LacZ*) was cotransformed with 400 µg pGBTCDC2A, 500 µg DNA of the library, and 40 mg salmon sperm carrier DNA using the lithium acetate method (Gietz *et al.*. 1992, Nucleic Acids Res. 20, 1425). To estimate the number of independent cotransformants, 1/1000 of the transformation mix was plated on Leu$^-$ and Trp$^-$ medium. The rest of the transformation mix was plated on medium to select for histidine prototrophy (Trp$^-$, Leu$^-$, His$^-$). Of a total of approximately 1.2 x 10^7 independent transformants 1200 colonies grew after 3 days of incubation at 30°C. The colonies larger than 2 mm were streaked on histidine-lacking medium supplemented with 10 mM 3-amino-1,2,4-triazole (Sigma). Two-hundred-fifty colonies capable of growing under these conditions were tested for β-galactosidase activity as described (Breedon and Nasmyth 1995, Cold Spring Harbor Symp. Quant. Biol. 50, p643-650), and 153 turned out to be His$^+$ and LacZ$^+$. Plasmid DNA was prepared from the positive clones and sequenced.

Example 3: Cell cycle interacting proteins associating with Cdc2aAt or Cdc2bAt

Nine cDNA clones were obtained by the method described in Example 1 and 2, which are further described below. The specificity of the interaction those clones was verified by the retransformation of yeast with pGBTCDC2A or pGBTCDC2B and the corresponding cDNA clones. As controls, pGBTCDC2A or pGBTCDC2B was cotransformed with a vector containing only the GAL4 activation domain (pGAD424); and the nine cDNA vectors were each cotransformed with a plasmid containing only the GAL4 DNA binding domain (pGBT9). Transformants were plated on medium with or without histidine. Only transformants containing both pGBTCDC2A or pGBTCDC2B and one of the nine cDNA clones were able to grow in the absence of histidine.

Example 4: Vb89 (SEQ ID NO: 7) – HAL3

BLAST analysis

A BLAST data base search revealed that the Vb89 clone encode the *Arabidopsis thaliana HAL3* homologue, isolated recently and of which the function was unknown. Unexpectingly, the Vb89 clone interacts with CDC2bAt, but not with CDC2aAt in the two-hybrid system. The interaction of Vb89 with CDC2bAt highlights an important role of Vb 89 in cell cycle control. The publicly available databases were screened with the cDNA VB89. An overall perfect homology with HAL3, already known in the databases was found. With the help of BLASTX U80192 (score 1.9e-106) was found as the best homologue. This sequence is a partial cDNA from A.thaliana (entered in the databank: 28-APR-1997)(with publ.: Culianez-Macia,F.A., Espinosa-Ruiz,A. and Serrano,R, Arabidopsis thaliana HAL3 homolog gene, unpublished). Except that VB89 is longer, there are no major differences with this cDNA.

HAL3 is a halotolerant gene isolated in *Saccharomyces cerevisiae* (Ferrando, 1995 Molecular and Cellular Biology, 15:5470-5481.). Hal3p can inhibit the Ppz1 protein phosphatase resulting in an increased resistance to sodium and lithium. These effect is largely a result of the increased expression of the *ENA/PMR2A* gene. This gene codes for a P-type ATPase responsible for sodium efflux (De Nadal et al., 1998 Proc. Natl. Acad. Sci. USA, 95: 7357-7362). The *HAL3* gene has also been isolated independently (as *SIS2*) and characterized on the basis of its ability to increase, when present in high copy number, the growth rate of *sit4* mutants (Di Como et al., 1995 Genetics, 139: 95-107.). The *SIT4/PPH1* gene encodes a type 2A-related Ser/Thr protein phosphatase that is required in late G1 for normal G1 cyclin expression and for bud formation. Interestingly, overexpression of *HAL3/SIS2* stimulates the rate of cyclin accumulation in *sit4* mutants.

Altering expression of gene

The Vb89 or HAL3 gene isolated according to the invention (and its homologs, derivatives and variants) may be used to confer salt tolerance on plants and/or improved growth under such conditions. The gene is expressed in plants using various types of

promoters, such as a constitutive promoter, a tissue-specific promoter, preferably a root-specific promoter or an inducible promoter, preferably a salt-inducible promoter.

Example 5: VbDAHP (SEQ ID NO: 9)

When a BLAST data base was used it was found that the VbDAHP clone encode a 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) synthase with a high similarity to the DHS2 gene. The VbDAHP clone interacts with CDC2bAt, but not with CDC2aAt in the two-hybrid system. The publicly available databases were screened with the cDNA VBDAHP (SEQ ID NO: 9). An overall perfect homology was found with DAHP (AROG_ARATH 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase [Arabidopsis thaliana]), already known in the databases. With the BLASTX as best homologue Q00218 (score 1.9e-49, C-term; 5.6e-86, N-term) was found. This sequence is a complete mRNA from A.thaliana (entered in the databank: 01-NOV-1997) (with publ.: Keith, Proc. Natl. Acad. Sci. U.S.A. 88 (19), 8821-8825 (1991)). With the BLASTN/nr we found the same DAHP.

In *Arabidopsis thaliana*, two genes has been isolated encoding 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) synthase, an enzyme catalyzing the first committed step in aromatic amino acid biosynthesis (Keith et al., 1991). Both genes, DHS1 and DHS2, may have distinct physiological roles, as there are differentially expressed in plants subjected either to physical wounding or to infiltration by virulent and avirulent strains of Pseudomonas syringae. Other enzymes in the *Arabidopsis* aromatic pathway are also encoded by duplicated genes, an arrangement that may allow independent regulation of aromatic amino acid biosynthesis by distinct physiological requirements such as protein synthesis and secondary metabolism.

Keith B., Dong X., Ausubel F.M., Fink G.R. (1991) Differential induction of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase genes in Arabidopsis thaliana by wounding and pathogenic attack. Proc. Natl. Acad. Sci. USA, 88: 8821-8825.

Example 6: VbHSF (SEQ ID NO: 13) – Heat Shock Factor 3

BLAST analysis

A BLAST data base search revealed that the VbHSF clone is very similar to the *Arabidopsis thaliana* Heat-Shock Transcription Factor HSF3. The VbHSF clone interacts with CDC2bAt, but not with CDC2aAt with the two-hybrid system. Organisms synthesize heat shock proteins (HSPs) in response to sublethal heat stress and concomitantly acquire increased tolerance against a subsequent, otherwise lethal, heat shock. Heat shock factor (HSF) is essential for the transcription of many HSP genes. Recently two HSF genes, HSF3 and HSF4, were isolated from an Arabidopsis cDNA library (Prandl et al., Mol Gen Genet (1998) May; 258(3):269-78). Transgenic *Arabidopsis* plants were generated containing constructs that allow expression of HSF3 and HSF4 or the respective translational beta-glucuronidase (GUS) fusions. Overexpression of HSF3 or HSF3-GUS, but not of HSF4 or HSF4-GUS, causes HSP synthesis at the non-heat-shock temperature of 25 degrees C in transgenic *Arabidopsis*. In transgenic plants bearing HSF3/HSF3-GUS, transcription of several heat shock genes is derepressed. Electrophoretic mobility shift assays suggest that derepression of the heat shock response is mediated by HSF3/HSF3-GUS functioning as transcription factor. HSF3/HSF3-GUS-overexpressing *Arabidopsis* plants show an increase in basal thermotolerance, indicating the importance of HSFs and HSF-regulated genes as determinants of thermoprotective processes. Plants transgenic for HSF3/HSF3-GUS exhibit no other obvious phenotypic alterations.

Derepression of HSF activity upon overexpression suggests the titration of a negative regulator of HSF3 or an intrinsic constitutive activity of HSF3. Stable overexpression of HSFs may be applied to other organisms as a means of derepressing the heat shock response.

A possible regulatory interaction between heat shock response and cell cycle control in plants has already been suggested. Reindl et al. (Plant Physiol (1997) Sep;115(1):93-100) reported the phosphorylation of the *Arabidopsis* heat-shock transcription factor HSF1 by a cyclin-dependent kinase. The HSF1 kinase forms a stable complex with AtHSF1, The HSF1 kinase interacts with the cell-cycle control protein Suc1p and is

immunoprecipitated by an antibody specific for the *Arabidopsis* cyclin-dependent CDC2a kinase. Phosphorylation by CDC2a in vitro inhibits DNA binding of AtHSF1 to the cognate heat-shock elements.

Different studies have shown that Heat shock factors can serve as auxillary proteins in formation of CDK/cyclin complexes. For example during meiosis I of mouse spermatocytes it is proposed that HSP70-2 assists in CDC2/cylinB1 complex formation through interaction with CDC2 and that this interaction establishes and/or maintains the CDC2 protein in a conformation that is competent for cyclin B1 binding (Zhu et al., 1997 Development 124: 3007-3014).

To obtain further independent evidence for the interaction of VbHSF with CDC2bAt, the VbHSF protein was overproduced in E.coli, purified to homogeneity, and coupled to Sepharose beads. The VbHSF-Sepharose beads were used during binding and kinase assays:

Expression and purification

For VbHSF expression and purification a fusion protein with a His-Tag sequence was generated. The VbHSF-coding region was PCR amplified using the primers 5'-CCATATGGAATTCGCACGAGGC-3' (SEQ ID NO: 21) and 5'-GCAGTAATAGGATCCACTATAGGG-3' (SEQ ID NO: 22). The PCR fragment was cut with NdeI and BamHI and cloned into the The NdeI and BamHI sites of pET19b (Novagen) and the resulting vector pETHSF was transformed into E.coli BL21 cells (Novagen). E.coli cells were grown at 37 °C until OD = 0.6 and the production of the fusion protein was induced by adding 1mM isopropyl-β-D-thiogalactopyranoside for 3 h at 30 °C. Cells were spun down and frozen. After thawing cells on ice, the cells were resuspended in lysis buffer containing 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl and 10 mM imidazole and lysozym was added to 1 mg/ml, After incubation on ice for 30 minutes, cells were sonicated and spun; the supernatant was loaded on Ni-NTA resin (Qiagen) and incubated for 1 h at 4 °C (200 rpm on a rotary shaker). Subsequnetly the lysate-Ni-NTA mixture was loaded on a column and the column was washed with 5 volumes of wash buffer containing 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl and 20 mM

imidazole. Next the fusion protein was eluted with 3 volumes elution buffer (50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl and 250 mM imidazole). The purified VbHSF protein was coupled to CNBr-activated Sepharose 4B (Pharmacia, Uppsala, Sweden) at a concentration of 5 mg/ml of gel according to the manufacturer's instruction.

Binding assay:

Protein extracts were prepared from 2-day-old cell suspensions of *A. thaliana* Col-O in homogenization buffer (HB) containing 50 mM Tris-HCL (pH 7.2), 60mM β-glycerophosphate, 15mM nitrophenyl phosphate, 15mM EGTA, 15mM $MgCl_2$, 2mM dithiothreitol, 0.1 mM vanadate, 50 mM NaF, 20 µg/ml leupeptin, 20 µg/ml aprotenin, 20 µg/ml soybean trypsin inhbitor (SBTI), 100 µM benzamidine, 1mM phenylmethylsulfonylfluoride, and 0.1 % Triton X-100. In a total volume of 300 µl HB, 150 mg of protein was loaded on 30 µl 50% (v/v) VbHSF-Sepharose or control Sepharose beads and incubated on a rotating wheel for 2h at 4 °C. Beads were washed 3 times with Beads Buffer containing 50 mM Tris-HCL (pH 7.2), 50 mM NaF, 250 mM NaCl, 5 mM EDTA, 0.1 % NP-40, 10 µg/ml leupeptin, 10 µg/ml aprotenin, 10 µg/ml SBTI, 100 µM benzamidine, 1mM phenylmethylsulfonylfluoride. Beads were resuspended in 25 µl SDS-loading buffer and boiled The suspernatant was separated on a 12.5 % SDS-PAGE gel and electroblotted on nitrocellulose membrane (Hybond-C^+; Amersham). Filters were blocked overnight with 2 % milk in phosphate buffered saline (PBS), washed 3 times with PBS, probed for 2 h with specific antibodies for CDC2aAt (1/5000 dilution) or CDC2bAt (1/2500 dilution) in PBS containing 0.5 % Tween-20 and 1 % albumin, washed for 1 h with PBS with peroxidase-conjugated secondary antibody (Amersham) and washed for 1 h with PBS containing 0.5 % Tween-20. Protein detection was done by the chemoluminescent procedure (Pierce, Rockford, IL). Using a CDC2bAt-specific antibody, no signal was observed at the expected size of CDC2bAt in extracts eluted from the control Sepharose beads. However a clear positive signal was observed for extracts loaded upon the VbHSF-Sepharose beads, giving independent evidence for the interaction between VbHSF and CDC2bAt.

Kinase assays:

The kinase assays were performed with Cdk complexes purified from total plant protein extracts by $p13^{suc1}$-sepharose affinity binding, according to Azzi et al. (1992). In order to control if the VbHSF protein has a regulatory effect on the phosphorylation of Histon H1, VbHSF-Sepharose and control beads were used during a Histon H1 kinase assay as described by Hemerly et al. (1995). After 20 min incubation at 30°C, samples were analysed by SDS-PAGE and autoradiographed. We could not detect any difference in [^{32}P] phosphate incorporation in histon H1 comparing the control and the VbHSF samples.

In another kinase assay we did not use Histon H1, to test wether the VbHSF functions as a substrate or not. The analysis of the autoradiography revealed the phosphorylation of the VbHSF protein by the CDK complexes. In combination with the binding assay we can speculate that the VbHSF protein acts as a substrate for CDC2bAt.

These results indicate that VbHSF (or HSF3) is phosphorylated by CDK. This suggests a regulatory role of phosphorylation of VbHSF by CDK/cyclin complexes namely that HSF3 activity is affected by phosphorylation, and hence its ability to confer thermotolerance on a plant may be manipulated.

Altering HSF3 expression to confer thermotolerance in plants

The invention provides a method for conferring thermotolerance on a plant by modifying the activity of HSF3, preferably via its phosphorylation state. Therefore, a nucleic acid of HSF3 is introduced into a plant cell, plant tissue or plant that encodes a HSF3 with a modified phosphorylation state.

It is possible to identify phosphorylation sites of HSF3 by random mutagenesis, anti-phospho-amino acid antibodies (e.g. anti-phospho-tyrosine and anti-phospho-threonine antibodies – Zhang. Planta 200 (1996), 2-12) and computer-assisted identification, methods known in the art. A state of enhanced phosphorylation can be mimicked by replacing the phosphorylated amino acids by a glutamic acid or aspartic acid. A method to prevent phosphorylation and to mimick a non phosphorylation state, comprises

replacing the phosphorylated amino acids by an amino acid that cannot be phosphorylated (other than glutamic acid or aspartic acid), namely an amino acid that is not tyrosine, serine or threonine. The invention would also relate to transgenic plants, tissues and cells obtainable by the methods above and comprising a HSF3 with modified activity. As mentioned previously such transgenic plants may display improved tolerance to stress, in particular heat stress.

Example 7: LDV24 (SEQ ID NO: 3) – PHO80-like protein (PLP)

BLAST analysis

The *LDV24* gene, renamed the PLP5 gene encodes a protein interacting with CDC2a and being highly similar to the PREG1 and PHO80 proteins of *Neurospora crassa* and *Saccharomyces cerevisiae*, respectively. The publicly available databases were screened with the cDNA LDV24. With the BLASTX as best homologue the PREG(AF051226) protein from Picea mariana (score: 1.5e-35) and PREG(AC003672) protein from Arabidopsis (score: 3.1e-35) were found. But there is homology with (P20052|PH80_YEAST) PHOSPHATE SYSTEM CYCLIN PHO80 (score: 2.1e-10). With the BLASTN/nr we found AF051226 Picea mariana PREG-like protein (score: 3.9e-12). Functional domains are predicted at amino acid positions 61-168 and 73-171 as comprising putative cyclin like interacting domains.

PHO80 itself shows similarity to the *Saccharomyces cerevisiae* G1-specific cyclins HCS26 and OrfD (Kaffman, Science 263 (1994) 1153-1155). The catalytic CDK subunit binding to PHO80 is PHO85, a CDK with roles in both the cell cycle and metabolic controls (Lenburg and O'Shea 1996, TIBS 21, p383-387). PHO80 in complex with PHO85 regulates phosphatase gene expression. When inorganic phosphate in the medium is abundant the PHO80-PHO85 complex phosphorylates the PHO4 transcription factor. Phosphorylated PHO4 remains mainly cytoplasmic, resulting in the repression of expression of the PHO5 phosphatase gene (O'Neill et al. 1996, Science 271, p209-212). When cell are starved for phosphate, the PHO80-PHO85 complex is inhibited by the CDK inhibitor PHO81, and transcription of PHO5 is activated.

The levels of PHO5 expression are sensitive to the levels of PHO80. Overexpression of PHO80 results in a partial defect of PHO5 activation when phosphate is limiting (Yoshida et al. 1989, MGG 217, p40-46; Madden et al. 1988, Nucleic Acids Res. 16, p2625-2637). At the other hand, deletion of PHO80 results in the presence of high levels of inorganic phosphate (Madden et al. 1988, Nucleic Acids Res. 16, p2625-2637). Similar effects can be expected for plants when the *LDV24* gene is deleted or overexpressed. This might result in an adapted growth in conditions where organic phosphate is present at limiting or exceeding levels. More phosphate accumulation might positively affect the rate of plant growth and biomass production.

Isolation of other PLPs

A systematic database screening using the PLP5 gene sequence as template revealed the existence of four related genes in *Arabidopsis thaliana* (see Table 1). These novel genes were isolated using the RT-PCR technology using the below enlisted combinations of primers. Briefly, total RNA was isolated from exponential phase cell suspension cultured of *Arabidopsis thaliana* ecotype columbia by the anidinium thiocyanate-phenol-chloroform method using an RNA extraction solution (TRIzol Reagent, GibcoBRL, Grand Island, NY). For the cDNA synthesis, was used the superscript preamplification system, taking 3 micrograms total RNA for the first strand synthesis using the oligo (dT) primer and follow the manufacturer's manual for the rest. 1 microliter of cDNA was used for isolating the five PLP genes by PCR using specific primers (see Table 2 and SEQ ID NOS: 23 to 32) and the following PCR reaction condition: initial denaturalisation 94 °C for 2 min, 35 cycles of 94 °C 45 sec, 55 °C 45 sec and 72 °C 1 min for each one, and a final extension at 72 °C for 5 min. After gel purification of all the cDNA the fragments were cloned directly into the vector pGEM-T and sequenced. The resulting nucleotide and amino acid sequences are given in Table 3a and SEQ ID NOS: 33 to 42. All proteins contain a highly conserved cyclin-like domain (see Table 3b and SEQ ID NOS: 43 to 47). Table 4 gives the percentage of sequence identity and similarity between the different PLP proteins.

Table 1: Database acknowledgements of P_L.

	EST	Annotation	GenBank accesion number	Chromosom
PLP1	T4E19	F16B22.1	O80513	II
PLP2	143B15T7,	-	-	V
PLP3	103D21XP, 316G7T7, 227G23T7, 103D21T7	-	-	III
PLP4	-	T14P1.11	AAD32828	II
PLP5	176E21T7, 213N15T7, 230B16T7, 138P12T7	-	-	

Arabidopsis thaliana WU-BLAST2 Search, Comparison Matrix : BLOSUM62

Table 2: List of the forward and reverse primers used for isolating the PLP1-4 genes.

	forward	reverse
PLP1	5'**GGGAATTC**ATGGCGGAACTTGAGAA TCC3'	5'**GGGGATCC**AAGACAAGATAAGAGTCCCTGC CG3'
PLP2	5'**GGGAATTC**ATGGCTGATCAGATTGA GATCC3'	5'**GGGGATCC**GCATAAATATAATCAAGCAGCA GCG3'
PLP3	5'**GGGAATTC**ATGTTAACCGCAGCCGG AGACG3'	5'**GGGGATCC**GGGGATCCATCAAACATATAAA GATG3'
PLP4	5'**CCGAATTC**ATGGATTCCCTAGCGATT TCTCC3'	5**GGGGATCC**CTACAACATGATTCGAGAAAATT GATGG3'
PLP5	5'**GGGAATTC**ATGGACTCTCTCGCAAC C3'	5'**GGGGATCC**TTGCCGATCAGCGTGC3'

Bold sequence: sequence included to facilitate cloning

Table 3a: cDNA and deduced amino-acid sequence of the *A. thaliana* PLP genes.

PLP1 (SEQ ID NOs: 33 and 34)

ATGGCGGAACTTGAGAATCCAAGTGTAATGTCGAAGCTGATAGCATTCTTATCTTCATTGCTA
GAGCGAGTTGCTGAGTCAAACGATCTGACCCGACGAGTCGCGACTCAGTCACAGAGAGTTTCG
GTGTTTCATGGACTGAGTCGACCAACGATAACGATTCAGAGCTATCTAGAGAGGATCTTCAAA
TACGCAAATTGTAGTCCTTCTTGCTTCGTCGTTGCTTACGTTTATCTCGATCGTTTCACTCAC
AGACAACCTTCACTTCCCATCAATTCCTTTAACGTCCATCGTCTTCTCATCACTAGTGTCATG
GTCGCTGCTAAATTCCTCGATGATCTGTACTACAACAATGCGTATTACGCGAAAGTGGGAGGA
ATAAGCACGAAGGAGATGAATTTTCTAGAGCTGGATTTCTTATTCGGGTTAGGATTTGAATTA
AACGTGACGCCAAACACATTCAACGCCTACTTCTCTTATCTTCAAAAGGAAATGACTCTTCTT
CAACCTCTCTCTCTCGTTGTTGTCCCATCATCAAGATCTCTCATTACCTTCAACGACGATGAA
GCTTCTCATCAGAAACAACAACAACAACAACTCGCTGTTTGA

MAELENPSVMSKLIAFLSSLLERVAESNDLTRRVATQSQRVSVFHGLSRPTITIQSYLERIFK
YANCSPSCFVVAYVYLDRFTHRQPSLPINSFNVHRLLITSVMVAAKFLDDLYYNNAYYAKVGG
ISTKEMNFLELDFLFGLGFELNVTPNTFNAYFSYLQKEMTLLQPLSLVVVPSSRSLITFNDDE
ASHQKQQQQQLAV

PLP2 (SEQ ID Nos: 35 and 36)

ATGGCTGATCAGATTGAGATCCAGAGAATGAACCAAGATCTTCAAGAACCATTGGCTGAGATC
ATGCCAAGTGTTTTAACGGCAATGTCGTATCTCTTGCAAAGAGTATCGGAGACCAACGACAAC
CTGAGCCAGAAACAGAAGCCCTCAAGCTTCACTGGAGTAACCAAACCTTCCATTTCCATCAGA
AGCTATCTCGAACGGATCTTTGAATACGCGAATTGTAGCTACTCGTGTTACATCGTCGCATAT
ATATATTTGGATCGGTTCGTGAAGAAGCAGCCATTTTTGCCTATCAATTCTTTTAATGTCCAT
AGGCTTATAATCACAAGTGTCTTGGTCTCTGCTAAATTCATGGATGACTTGAGTTACAACAAT
GAATATTATGCAAAAGTTGGAGGAATAAGCAGAGAAGAAATGAACATGCTTGAGCTTGACTTC
TTGTTCGGAATTGGGTTTGAGTTAAACGTCACCGTTTCTACTTTCAATAACTATTGTTGTTTT
CTACAAAGAGAGATGGCGATGTTGATGAAGATGAAGTCTCTGTTTCTTGAACCTTCTTCATTC
AAAATCTCTTTTAAGACGAAACTTGTGATGTATCCACACGAGGAAGACTCTTTATCTACTCAC
CACAACAAGAAGCAACTCGCTGCTGCTTGA

MADQIEIQRMNQDLQEPLAEIMPSVLTAMSYLLQRVSETNDNLSQKQKPSSFTGVTKPSISIR
SYLERIFEYANCSYSCYIVAYIYLDRFVKKQPFLPINSFNVHRLIITSVLVSAKFMDDLSYNN
EYYAKVGGISREEMNMLELDFLFGIGFELNVTVSTFNNYCCFLQREMAMLMKMKSLFLEPSSF
KISFKTKLVMYPHEEDSLSTHHNKKQLAAA*

PLP3 (SEQ ID Nos: 37 and 38)

PLP3 (SEQ ID Nos: 37 and 38)

ATGTTAACCGCAGCCGGAGACGATGAACTGGACCCGGTCGTGGGACCAGAATCGGCAACGGAA
GCAGCCACTCCAAGAGTGCTGACTATAATCTCCCATGTGATGGAGAAGCTCGTGGCACGAAAC
GAGTGGTTAGCTAAGCAAACTAAGGGATTTGGGAAGAGCTTGGAGGCGTTTCACGGCGTGAGA
GCGCCGAGCATAAGTATAGCTAAATACCTTGAGAGGATATATAAGTACACAAAATGTAGCCCG
GCATGTTTCGTTGTTGGGTATGTGTACATAGACCGGTTGGCTCATAAGCATCCTGGTTCTTTG
GTTGTCTCCTTGAATGTTCATAGACTCCTCGTCACTTGTGTCATGATTGCTGCCAAGATACTA
GATGACGTGCACTACAACAACGAGTTCTATGCTCGGGTTGGAGGCGTAAGCAATGCAGACTTG
AACAAAATGGAGTTGGAGCTTCTCTTTCTTCTTGACTTTAGAGTTACTGTGAGTTTTAGAGTT
TTCGAGAGCTATTGCTTTCACCTCGAAAAAGAGATGCAACTAAACGACGTCGTTTCTTCCCTC
AAAGATATTCAACCAATGCAAGAAAGTCTCTCTCCAGCATCTACTTTATCATCTTTATATGTT
TGA

MLTAAGDDELDPVVGPESATEAATPRVLTIISHVMEKLVARNEWLAKQTKGFGKSLEAFHGVR
APSISIAKYLERIYKYTKCSPACFVVGYVYIDRLAHKHPGSLVVSLNVHRLLVTCVMIAAKIL
DDVHYNNEFYARVGGVSNADLNKMELELLFLLDFRVTVSFRVFESYCFHLEKEMQLNDVVSSL
KDIQPMQESLSPASTLSSLYV

PLP4 (SEQ ID Nos: 39 and 40)

ATGGATTCCCTAGCGATTTCTCCAAGGAAGCTCCGATCAGACCTCTACTCTTACTCTTACCAA
GATGATTCCAACACAGTACCTCTAGTCATCTCTGTTCTCTCGTCTCTGATCGAACGAACTTTA

GCTAGGAACGAGAGAATCAGCCGGAGCTACGGTGGTTTTGGTAAGACACGTGTCTTTGATTGC
CGGGAGATTCCTGATATGACTATTCAATCATACCTAGAGAGAATTTTCCGGTATACCAAAGCC
GGTCCATCGGTTTACGTCGTGGCTTATGTATACATTGACCGGTTCTGTCAGAATAACCAAGGT
TTCAGAATCAGTCTTACCAATGTACATCGTCTCCTTATCACAACTATCATGATCGCTTCCAAA
TACGTCGAAGATATGAACTACAAAAACTCGTACTTTGCGAAAGTAGGAGGATTAGAGACAGAA
GATTTGAACAATTTGGAACTGGAGTTCTTGTTCTTGATGGGATTTAAGTTGCATGTGAATGTG
AGTGTGTTCGAGAGTTACTGTTGTCATCTAGAAAGAGAAGTGAGTATTGGAGGAGGTTATCAG
ATCGAAAAAGCATTGCGTTGCGCTGAGGAAATCAAATCTAGACAAATTGTTCAAGACCCTAAA
CATCATCATCACCATCAATTTTCTCGAATCATGTTGTAG

MDSLAISPRKLRSDLYSYSYQDDSNTVPLVISVLSSLIERTLARNERISRSYGGFGKTRVFDC

REIPDMTIQSYLERIFRYTKAGPSVYVVAYVYIDRFCQNNQGFRISLTNVHRLLITTIMIASK

YVEDMNYKNSYFAKVGGLETEDLNNLELEFLFLMGFKLHVNVSVFESYCCHLEREVSIGGGYQ

IEKALRCAEEIKSRQIVQDPKHHHHHQFSRIML

PLP5 (SEQ ID Nos: 41 and 42)

ATGGACTCTCTCGCAACCGATCCAGCTTTCATTGATTCGGATGTATACCTCAGGTTAGGACTT
ATTATTGAGGGCAAACGATTGAAAAAGCCACCGACTGTACTCTCACGCCTCTCTTCTTCTCTG
GAGAGATCTCTGTTACTCAATCATGATGACAAGATTCTGCTTGGATCGCCAGACTCTGTTACC
GTGTTTGACGGGAGATCTCCCCCTGAGATCAGTATTGCACACTACTTGGATCGCATTTTCAAG
TACTCTTGCTGCAGTCCCTCCTGCTTCGTCATTGCGCATATCTACATTGATCACTTTCTCCAT
AAGACCCGAGCCCTTCTCAAACCCCTTAATGTCCACCGCCTTATCATTACAACTGTCATGTTA
GCTGCTAAAGTCTTCGATGATAGGTATTTCAACAATGCATACTACGCAAGAGTGGGAGGTGTG
ACTACGAGAGAGTTAAACAGATTGGAGATGGAGTTGTTGTTTACCCTTGACTTCAAGCTTCAG
GTAGATCCTCAGACGTTTCACACACACTGTTGTCAGTTAGAAAAGCAGAACAGAGACGGCTTC
CAGATCGAGTGGCCCATAAAAGAAGCATGCCGAGCCAACAAAGAGACTTGGCAGAAGAGGACA
CCCGACTCACTCTGCTCTCAAACCACAGCACGCTGATCGGC

MDSLATDPAFIDSDVYLRLGLIIEGKRLKKPPTVLSRLSSSLERSLLLNHDDKILLGSPDSVT

VFDGRSPPEISIAHYLDRIFKYSCCSPSCFVIAHIYIDHFLHKTRALLKPLNVHRLIITTVML

AAKVFDDRYFNNAYYARVGGVTTRELNRLEMELLFTLDFKLQVDPQTFHTHCCQLEKQNRDGF

QIEWPIKEACRANKETWQKRTPDSLCSQTTAR

Table 3b: Domains analysis

	Position	Amino Acid Sequence	CYCLIN domain[1]
PLP1	56 – 141 (SEQ ID NO: 43)	YLERIFKYANCSPSCFVVAYVYLDRFTHRQPSLPINSF NVHRLLITSVMVAAKFLDDLYYNNAYYAKVGGISTKEM NFLELDFLF	1.24955
PLP2	64 – 149 (SEQ ID NO: 44)	YLERIFEYANCSYSCYIVAYIYLDRFVKKQPFLPINSF NVHRLIITSVLVSAKFMDDLSYNNEYYAKVGGISREEM NMLELDFLF	1.19454
PLP3	71 –156 (SEQ ID NO: 45)	YLERIYKYTKCSPACFVVGYVYIDRLAHKHPGSLVVSL NVHRLLVTCVMIAAKILDDVHYNNEFYARVGGVSNADL NKMELELLF	0.63602
PLP4	73 – 158 (SEQ ID NO: 46)	YLERIFRYTKAGPSVYVVAYVYIDRFCQNNQGFRISLT NVHRLLITTIMIASKYVEDMNYKNSYFAKVGGLETEDL NNLELEFLF	1.87380
PLP5	77 – 161 (SEQ ID NO: 47)	YLDRIFKYSCCSPSCFVIAHIYIDHFLHKTRALLKPLN VHRLIITTVMLAAKVFDDRYFNNAYYARVGGVTTRELN RLEMELLF	0.19729

Notes: Domain present in cyclins, TFIIB and Retinoblastoma
[1] E-values are calculated using Hidden Markov Models.

Table 4: Amino acid sequence identity and similarity (bold) between the different *A. thaliana* PLPs.

	PLP1	PLP2	PLP3	PLP4	PLP5
PLP1	-	61	45	42	48
PLP2	**80**	-	41	36	44
PLP3	**68**	**67**	-	45	35
PLP4	**68**	**62**	**68**	-	41
PLP5	**66**	**63**	**64**	**60**	-

Expression analysis of PLP genes in plants

The spatial expression pattern of the different PLP genes was studied using quantitative RT-PCR using the Superscript preamplification system (Gibco/BRL, Gaithersburg, MD, USA). Total RNA was isolated from roots, rosette leaves, stems, flowers, seedlings, and actively dividing cell suspensions using the Trizol reagents according to the manufacturer's protocol. First strand cDNA was synthesised from 1 microgram of RNA as described by the manufacturer. The single-stranded cDNA products were subjected to PCR using 0.2 mM concentrations of 5' and 3' specific primers (see Table 2). Care was taken to quantify changes in individual mRNA levels by employing appropriate RT-PCR conditions under which a linear relationship existed between amounts of RNA added and intensities of the RT-PCR products. Aliquots of 10 microliter were taken after the 15, 20 and 25 cycles, each cycle being 94°C for 30 s, 55°C for 30 s, and 72 °C for 1 minute. The products were electrophorically separated on a 1.0% agarose gel, stained with ethidium bromide and blotted onto nitrocellulose membranes. Fluorescein labelled probes specific for the different PLP genes were prepared using the Gene images random prime labelling module (Amersham). Signals were visualised using the Genes images CDP-star detection module (Amersham). A hybridising signal for PLP3 could only be observed for root tissue. In contrast PLP1, PLP2, PLP4, and PLP5 gene expression could be detected in all tissues examined, see Figure 1.

Interaction between the PLP and CDKs

Protein-protein interactions between the different PLPs and CDKs were studied using a two-hybrid system based upon GAL4 recognition sites to regulate the expression of the his3 reporter gene. Vectors and strains used were provide with the Matchmaker Two-Hybrid (Clontech, Palo Alto, CA). The baits used for the two-hybrid analysis were constructed by inserting the PLPs coding region into the pGBT9 (as an fusion protein with the DNA binding domain of the GAL4 transcription factor) and pGAD424 (as an fusion protein with the transcriptional activation domain of the GAL4 transcription factor) vectors. The inserts were created by PCR using the PLPs cDNA as template and primers to incorporate *Eco*RI and *Bam*HI restriction enzyme sites (see Table 2), resulting into the plasmids pGTBPLP1 to pGBTPLP5 and pGADPLP1 to pGADPLP5. Vectors were tested for self activation, and pGBTPLP2, pGBTPLP3 and pGBTPLP5 were found positive, excluding their use for studying protein-protein interactions. All

other constructions were tested for their interaction with the CDC2aAt and CDC2bAt proteins, cloned in pGTB9 and pGAD424. (De Veylder et al. (1997) *FEBS Lett* 412, 446-52). For this an appropriate reporter strain (HF7c (*MAT*$_a$ *ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3,112 gal4-542 gal80-538 LYS2::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-HIS3 URA3::GAL4*$_{17mers(3x)}$*-CyC1*$_{TATA}$*-LacZ*)) was transformed with different combinations of the two-hybrid vectors, and tested for its ability to grow in the absence of histidine. The obtained results are summarised in Table 5. All PLPs were shown to interact with CDC2aAt. PLP2 and PLP3 interact only with CDC2aAt, not with CDC2bAt. In contrast, PLP1, PLP4, and PLP5 interact with both CDC2aAt and CDC2bAt, but stronger with CDC2aAt.

Table 5: Two hybrid interaction between the PLPs and CDC2a and CDC2b genes.

pGBT9	pGADCDC2a	pGADCDC2b	pGAD424 (control)
PLP1	+++	+	-
PLP2	ND	ND	+++
PLP3	ND	ND	+++
PLP4	++	+	-
PLP5	ND	ND	+++

PGAD424	pGBTCDC2a	pGBTCDC2b	PGBT9 (control)
PLP1	+++	+	-
PLP2	+++	-	-
PLP3	+++	-	-
PLP4	++	+	-
PLP5	+++	+	-

Note: + interaction, - no interaction, ND no determinate

Isolation of PLP5 (cyclin PHO80) Arabidopsis mutant

In plant, a direct way for obtaining information on the function of a gene of interest is to study the gene disrupted mutant plant (Reverse genetics).

To identify a mutant plant, DNA extracted from pools of a collection of mutagenized plants generated for example by the insertion of a T-DNA element,are used as template for PCR screening using oligonucleotide primers from the insertional element and from the gene of interest. The sensitivity of the PCR reaction is able to detect the insertion of a T-DNA in the target gene. Once a pool has been confirmed to contain the interest

gene linked to the insertional element, the different mutant plants used to prepare the pool are analysed by PCR in order to identify the individual mutant line.

A. Identification of pools containing T-DNA insertion mutant in PLP5

1. Arabidopsis T-DNA insertion mutant collection :

At INRA-Versailles, a large population of mutagenized Arabidopsis plants, ecotype Wassilevskija (WS), has been generated by a vacuum and detergent infiltration methods (Becthold et al., 1993 ; 1995) with an Agrobacterium suspension strain MP5-1 carrying the binary vector pGKB5 (Bouchez et al., 1993, Becthold et al., 1995). At present, the collection contain more than 35,000 independent T-DNA lines with more of 55,000 inserts. (an average of one insertion every 2.5 kb).

For reverse genetics screens, the seeds of the generated T-DNA lines are grouped in primary pools of 48 families. Approximately 100 seeds from each family are mixed and ground in vitro on a large Petri plate. The seedlings plants (10-15 days, stage 2 rossetes leaves) are used for DNA extraction as described Doyle and Doyle, Focus 12:13-15.

Aliquots of 20 ul of the resuspended DNA (100-300 ng/ul) from each of the 16 primary pools are used to prepare 2 ml of one hyper-pools . Each hyper-pool represents 768 independent T-DNA lines. Aliquots of 5 ul (15-30 ng/ul) of each hyper-pools (46 hyper-pools, at present) are charged in a 96-well microplate, where the PCR amplification reaction will be performed.

References

Becthold N., Ellis J., and Pelletier G. (1993).In planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. C.R. Acad. Sci Paris, Sciences de la vie /Life Sciences 316 :1194-1119

Bouchez D., Camilleri C., Caboche M. (1993). A binary vector based on Basta resistance for in planta transformation of Arabidopsis thaliana. C.R. Acad. Sci. Paris, Sci de la vie/ Life Sciences 316 :1188-1193.

Becthold N., Bouchez D. (1995). In planta Agrobacterium-mediated transformation of adult plant Arabidopsis thaliana plants by vacuum infiltration. In : Gene transfer to plants. I. Potrykus and G. Spangenberg Eds, Springer-Verlag, Heilderberg, pp 19-23.

WEB pages :

T-DNA lines : http://nasc.nott.ac.uk : 8300/Vol2ii/pelletier.html

PGKB5 sequence : http://nasc.nott.ac.uk :8300 /Vol2ii/bouchez.html

2. PCR screening :

The oligonucleotides primers for the Arabidopsis cyc PHO were designed from the cDNA sequence obtained from the identified clone interacting with the Arabidopsis cdc2a kinase in a two hybrid screen. A foward and reverse primers were tested for specificity, and yield a good PCR amplification using the wild-type genomic DNA of Arabidopisis plants, ecotype WS, as template. The designed primers did not show unespecific amplification in combination with the T-DNA primers Tag3 nor Tag5.

Primers

Foward primer F2 : 5'-ATTGCACACTACTTGGATCGCATT-3' (SEQ ID NO: 48)

Reverse primer R1 :5'-GATAGAATGGGAACGGCTAG-3' (SEQ ID NO: 49)

Tag3 primer :5'-CTGATACCAGACGTTGCCCGCATAA-3' (SEQ ID NO: 50)

Tag5 primer :5'-CTACAAATTGCCTTTTCTTATCGAC-3' (SEQ ID NO: 51)

Each gene primer was used in combination with a T-DNA primer, in the PCR screen.

Standard PCR mix for each microplate well:

ADN :	5 ul (10-30 ng /ul)
PCR buffer :	2.5 ul
$MgCl_2$:	2.5 ul (25 mM)
dNTPs :	0.5 ul (10 mM)
Gene primer :	2.5 ul (10 uM)
T-DNA primer :	2.5 ul (10 uM)
Taq Polymerase :	1.0 ul (1U /ul)
H_2O :	8.5 ul

The PCR conditions were:

2' –94 C

---

10 cycles (touch down)

15"-94 C

30"–65 C -1 C /cycle

2' –72 C

---

35 cycles

15"-94 C

15"–55 C

1' –72 C

2' –72 C

5' –4 C

---

3. Hybridization Analysis

Due to the numerous artifacts generated by the PCR reaction, it is neccessary to identify, between the PCR products which one contain the gene of interest linked to the insertion element. To overcome this problem, an hybridization analysis was carried out.

5 ul of each hyper-pool PCR reaction were electrophored on a 2% agarose gel (TAE). After migration, the gel was equilibrated in 0.4 N of NaOH for 30 min, and transfered simultaneously to two charged nylon membranes (Pall, plus) over night. After transfer, the membranes were rinsed with 2X SSC, one followed by hybridization with the gene probe and the other with the T-DNA probes.

The gene probe was prepared from the digested plasmid containing the cDNA encoding for the cyclin PHO 80 identifed in the two hybrid screen. The T-DNA probes correspond to a mix of left border (fragment of 1kb after KpnI digestion of plasmid pBS-LB) and right border (fragment of 0,8 kb after SstI-EcoRV digestion of the pBS-RB)

The digested gel purifed fragments were labelled using the ALKPHOS (Amersham, RPN 3680) non-radiactive labelling kit. The hybridization and washing were done according with the instructions of the manufacturer.

Developing of each autoradiogram obtained after hybridization with the gene and the T-DNA probes, revealed a clear signal that superimposed in both blots, indicating a potential T-DNA insertion mutant. The PCR fragment given the hybridization positive signal was further sequenced confirming that it contained the cyclin PLP5 gene linked to the T-DNA insertion element.

The sequence of the mutant line was done with the forward primer 2 of cyclin PLP5.

Sequence of the mutant line with forward F2 primer:

NTGTACTAAA AGGTGCANTC CCTCCTGCTT CGTCATGGAT ATCTACATTG ATCACTTTCT
CCATAAGACC CGAGCCCTTC TCAAACCCCT TAATGTCCAC CGCCTTATCA TTACAACTGT
CATGTTAGCT GCTAAAGTCT TCGATGATAG GTATGTTACT CACTAAACCT GGTATCAAAT
TCAACACGCA AATAAGTCTT CAATCATAGA TTCATTGATC TCTGGTGTTG NGCAGGTATT
TCAACAATGC ATACTACGCA AGAGTGGGAG GTGTGACTAC GAGAGAGTTA AACAGATTGG
AGATGGAGTT GTTGTTTACC CTTGACTTCA AGCTTCAGGT AGATCCTCAG ACGTTTCACA
CACACTG**TTG TACTGAATCG GATTTTCAAG GGTCTGGCCA AAACTATTCC GNGGGCACCT**
GGCACACGCC CTGGAGTCCG GCCCGTTTCC AGTTGAGGGT TGTCTACGCT TANATGAGAA
GGAAAGTTGT CCAANACGAA TCCCAGTGTC CTATTACCAA TAGCCGACGG TATCGATAAG
CTNGATGTAC ATGGTCNATA NNAAAAGGCN AT (SEQ ID NO: 52)

Sequence homology to the right border of the T-DNA is indicated in bold.

The gene sequence is homologous to the Arabidopsis EST N37922, however there is not a genomic sequence homolog in the data base. At the protein level, it is homologous (score of 5e-30) to the PREG-like protein of Arabidopsis (AC003672), to the yeast cyclin PCL 7 partner of the cdc PHO 85 (score 1e-9), and to the yeast cyclin PHO80 (score 1e-6).

The length of the Arabidopsis PREG-like protein (AC003672) homologous to the cyclin PHO 80, is 202 aminoacids. If the PLP5 belong to the family of the PREG-like proteins, the T-DNA insertion should be located approximaly at aminoacid position 157.

4. Partial genomic sequence

Given that there is not genomic sequence homologs to the cDNAs of PLP5 in the data base, two oligonucleotids primers designed from the cDNA of PLP5 previously identifed

in the two hybrid screen, were used to amplify by PCR, a partial genomic fragment containing the corresponding cDNA sequence.

Forward primer F1 : 5'-cgatccagctttcattgattcg-3' (SEQ ID NO: 53)
Reverse primer R1 : 5'-GATAGAATGGGAACGGCTAG-3' (SEQ ID NO: 54)

The PCR fragment obtained was sequenced by dye terminator using the forward primer F1.

The sequence is :

ATTTCNTTNGNTGTATACCTCAGGTTAGGACTTATTATTGAGGGCAAACGATTGAAAAAGCCACC
GACTGTTCTCTCACGCCTCTCTTCTTCTCTGGAGAGATCTCTGTTACTCAATCATGATGACAAGA
TTCTGCTTGGATCGCCAGACTCTGTTACCGTGTTTGACGGGAGATCTCCCCCTGAGATCAGTATT
GCACACTACTTGGATCGCATTTTCAAGTACTCTTGCTGCAGTCCCTCCTGCTTCGTCATTGCGCA
TATCTACATTGATCACTTTCTCCATAAGACCCGAGCCCTTCTCAAACCCCTTAATGTCCACCGCC
TTATCATTACAACTGTCATGTTAGCTGCTAAAGTCTTCGATGATAGGTAT**GTTACTCACTAAACC**
TGGTATCAAATTCAACACGCAAATAAGTCTTCAATCATAGATTCATTGATCTCTGGTGTTGTGCA
GGTATTTCAACAATGCATACTACGCAAGAGTGGGAGGTGTGACTACGAGAGAGTTAAACAGATTG
GAGATGGAGTTGTTGTTTACCCTTGACTTCAAGCTTCAGGTAGATCCTCAGACGTTTCACACACA
CTGTTGTCAAGTTAGAAAAGCAGAACAGCGACGGCTTCCAGATCGAGTGGCCCATAAAAGAAGCA
TGCCGAGCCAACAAAGAGACTTGGCAGAAGAGGACACCCGACTCACTCTGCTCTCAAACCACAGC
ACGCTGATCGGCAAGGGNAAAANGA (SEQ ID NO: 55)

The alignment of the cDNA with the partial genomic sequence revealed the presence of one intron indicated in bold. The underlined sequece represent the insertion site of the T-DNA in the mutant line.

B. Identification of lines containing T-DNA insertion mutant in cyclin PHO80

Experimental design:

-identify a PLP5 insertion mutant
-characterize the PLP5 mutant
-identify homozygous plants
-select growth conditions to detect phenotype differences between wild type control and PLP5 mutants

1. Identification of the positive line

The 48 lines from the positive pool were grown in growth chamber for two weeks. Plants were harvested and frozen in liquid nitrogen. Plants (1g) were grinded with a pestle and a mortar and homogeneized in 6ml buffer containing 78mM Tris HCl pH8, 40mM EDTA, 390 mM NaCl, 1% SDS, 15mM sodium bisulfite at 65°C for 30 min. 2ml potassium acetate 5M were added and the mixture was incubated on ice for 20 min. Supernatants were recovered after centrifugation (20 min, 4500 rpm, 4°C) and 4ml isopropanol (-20°C) was added and incubated for at least 30 min at 4°C. After centrifugation (7 min, 4500 rpm, 4°C) the pellet was dried and taken in 420 microliter ammonium acetate 7.4M. Supernatants were recovered after centrifugation (20 min, 4500 rpm, 4°C) and 700 microliter isopropanol were added and incubated for 30 min at 4°C. After centrifugation (7 min, 4500 rpm, 4°C) the pellet was dried and taken in 400 microliters Tris-EDTA (100mM-10mM) buffer pH 8.0. After centrifugation (15 min, 13700 rpm, 4°C), supernatants were mixed with 800 microliters ethanol at -20°C for 10 min. The final pellet was recoved after centrifugation (5min, 13700 rpm, 4°C) and washed with 70% ethanol. The final pellet was taken in 20 microliters of Tris-EDTA (100mM-10mM) buffer pH8.0 and used for PCR (1/100 dilution).

2. Growth of positive lines

The positive line identified from the INRA collection was grown in growth chamber under four types of conditions:

1- 100mg/L kanamycin on At medium (see 6. General methods - below)

2- At medium minus sugar and - vitamins

3- At medium in light conditions (20 °C, 12 h photoperiod, normal intensity)

4- At medium in dark conditions

Plants were then examined for obvious phenotypes and kanamycin segregation which gives an indication on the number of T-DNA insertions, the linkage of insertions and the sex effect.

For other phenotypes of interest plants are grown on specific medium:

1-At medium minus sucrose plus different amounts of Pi (0 to 50mM K2PO4)

2-At medium plus different amounts of Pi (0 to 50mM)

3-At medium minus sucrose plus different amounts of Pi (0 to 50mM) at 28°C
4-At medium plus various amount of hygromycin (0 to 200mM)
5- At medium plus various amount of auxin, or cytokinines

Observations were made concerning germination, emergence of radicle, emergence of cotyledons, emergence of first pair of leaves, color, shape. Flowers were observed in green house on the homozygous lines.

3. Detection of Homozygous plants

Homozygous plants were detected by PCR first using the following combination of primers (F2-Tag5 and F2-R1).
F2: 5' ATTGCACACTACTTGGATCGCATT 3' (SEQ ID NO: 56)
R1: 5' CTATCTTACCCTTGCCGATCAGC 3' (SEQ ID NO: 57)
Tag5: 5' CTACAAATTGCCTTTTCTTATCGAC 3' (SEQ ID NO: 58)

PCR conditions were for one reaction:
5 microliter DNA or controle (water, wild type, pool)
2,5 microliter buffer (Tris-HCl 100mM pH9.5, KCl 500mM, 1% Triton X100)
2,5 microliter MgCl2 25mM
0.5 microliter dNucleotidesTP
1 microliter of TaqPolymerase (1Unit/microliter)
8.5 microliter water
2.5 microliter of each primer

The PCR program was as follow:
2 min 94°C
10 cycles touch-down 15 sec 94°C, 30 sec 65°C, 2 min 72°C
35 cycles 15 sec 94°C, 15 sec 55°C, 1 min 72°C
2 min 72 °C
5 min 4°C

Young leaves (1cm square) were ground in homogeneization medium (200mM Tris-HCl pH7.5, 250mM NaCl, 25mM EDTA, 0.5% sodium dodecyl sulfate). After 30 min on ice, supernatant was recovered after centrifugation (5min, 13000rpm; room temperature) and 1 volume isopropanol was added. After inversion of the tube and about 10 min, pellets were carefully recovered after centrifugation (5min, 13000rpm; room temperature), dried and taken into 20 microliter Tris-EDTA (100mM-10mM) buffer. PCR were done using 1:100 dilution of the DNA of each individual plants.

These plants were then transfered to the green house for multiplication and crosses. The seeds were then harvested and put in growth chamber on agarose plates containing 100mg/l kanamycin to analyse the segregation of the insertion into the PLP5 gene.

4. Growth of plants

In addition plants were grown directly in the green house on soil, watered, under 12 h photoperiod and normal light intensity. Such plants were also used to make crosses with wild type plants in order to clean the genotype from unwanted short T-DNA insertions in other genes not detected by kanamycin resistance gene.

5. Determination of GUS activity.

Gus activity is expressed when a T-DNA is inserted into a gene in the proper direction; This allows to detect where mutated gene is expressed. Preliminary data (Nusseaume, CEA Cadarache) indicated that GUS activity could be detected in the positive line containing PLP5 mutant.

Gus activity was detected as in Jefferson et al.., 1987 (EMBO J, 6: 3901-3907) with slight modifications. Tissues (whole plantlets: two weeks old) were fixed in 80%acetone for 1h at -20°C. Tissues were then incubated with 1mg/ml 5-bromo,4-chloro, 3indolyl, beta Dglucuronide in 0.1M potassium phosphate buffer pH7 containing 0.1% triton X100, 10mM EDTA, 2mM potassium ferrocyanide, 2mM potassium ferricyanide. Tissues were vacuum infiltrated for 10 min and incubated for at least 1 hour at 37°C. 70% ethanol was used to destain the tissues prior to microscopic analysis.

6. General methods/protocols/materials

Protocol for Arabidopsis culture:

• **seed sterilization** :

- Dissolve 1 bleach pellet (Bayrochlore, Bayrol) in 40 ml H_2O, add a few drops of Teepol or Tween (prepare fresh).

- Dilute the previous solution 1/10 in 95% Ethanol to make the sterilization solution (SS).

- Add 10 ml of SS to each tube of seeds, and incubate 7 min at room temperature with constant, gentle agitation.

- Rince twice with 10 ml 95% ethanol.

- Let the seeds sediment, and carefully remove as much ethanol as possible. (It's possible to invert the tubes, but be careful !).

- Leave the tubes open under a sterile hood overnight.

• ***in vitro* culture medium:**

We use non diluted medium with 10 g/l sucrose . (modified from Estelle & Sommerville, 1987, MGG 206 : 200).

At medium :

	Stock	Amount/L	Final concentration
KNO_3	1 M	5 mL	5 mM
KH_2PO_4	1 M	2,5 mL	2.5mM
$MgSO_4$	1 M	2 mL	2 mM
$Ca(NO_3)_2$	1 M	2 mL	2 mM
Microelements	1000x	1 mL	1x
Vitamins	500x	2 mL	1x
Bromcresol Purple	0.16%	5 mL	0.0008%
MES pH 6	14%	5 mL	0.035%
Agar		7 g	0.7%
Autoclave (120°C, 20 min), then add :			
Ferric Ammonium Citrate (autoclaved separately)	1%	5 mL	0.005%

• Sprinkle the surface-sterilized seeds on a 14 cm agar plate containing *Arabidopsis* culture medium (AtM/2), covered with a round filter paper (Whatman 3MM). Seal the plates with a gas-permeable chirurgical tape.

• Synchronize germination by a cold treatment at 4°C for 48 hours.

• Place in the growth chamber under the following conditions : photoperiod 16 h day (100-150 $\mu E/m^2/s$) / 8 h night ; temperature 20°C day / 15°C night ; humidity 70%.

• After 10-15 days of culture, plantlets (2-leaf rosettes) are ready for DNA isolation. Each plate should yield 3-6 g fresh weight.

• Gently scrape the plantlets from the filter paper using a razor blade. The plantlets are weighted, and frozen in liquid nitrogen for future use.

Microelements 1000x :

	Amount/L (1000 x)	Final concentration (1x)
H_3BO_3	4328 mg	70 μM
$MnCl_2$, $4H_20$	2770 mg	14 μM
$CuSO_4$, $5H_20$	125 mg	0.5 μM
Na_2MoO_4, $2H_20$	50 mg	0.2 μM
NaCl	584 mg	10 μM
$ZnSO_4$, $7H_20$	288 mg	1 μM
$CoCl_2$, $6H_20$	2.5 mg	0.01 μM

Autoclave (120°C, 20 min)

Vitamins 500x :

	Amount/L (500 x)
Myo-Inositol	50 g/L
Ca Panthotenate	0.5 g/L
Niacin	0.5 g/L
Pyridoxine	0.5 g/L
Thiamine HCl	0.5 g/L

Biotin 5 mg/L

Keep at -20°C.

Using the above methods a mutant containing an insertion in the PLP5 gene was identified and called Mutant 11 (mut11).

(a) Segregation analysis for kanamycin resistance indicated a ¼/¾ population of sensitive/resistant plants. This indicates the probability for one insertion of T-DNA. This result was confirmed by southern analysis of DNA digestions with T-DNA probes.

(b) Effects of hygromycin: Aminoglycosides are antibiotics that affect rRNA interactions and lead to mistranslation. It was shown that in yeast pho80-pho85 and pho4 are required for increased sensitivity to aminoglycoside antibiotics (Wickert et al. (1998) J. Bacteriology 180 (7):1887-1894). Mut11 is tested to determine whether it is hypersensitive to hygromycin. Plants were grown on At medium plus various amounts of hygromycin. Observations were made for germination, cotyledon emergence, and general root aspect.

Conclusions:

- at low concentrations final germination capacity is similar for WS and 11K11 but mean time germination is longer (about 18h) for homozygous mut11 at 10mM. A similar phenomenon is observed at 25mM hygromycin but final germination is similar. As a consequence cotyledon emergence is delayed.
- at higher concentrations, mean time germination is longer for mut11 and final germination capacity is reduced. Cotyledon emergence is delayed for 11K11 and radicle growth is severely affected for both wild type and 11K11.
- as a conclusion, mut11 is more sensitive to hygromycin, suggesting a role of pho80 and/or other components of the signalling cascade in sensitivity to hygromycin. Transgenics overexpressing mut11 could be more resistant to hygromycin. Overexpression of the PLP5 would mean that could be used as a positive selectable marker during transformation procedures while antisense/cosuppression could be used as a negative selective marker.

Example 8: Extension of cell cycle interacting protein encoding polynucleotides to full length or to recover regulatory elements

The cell cycle interacting protein encoding nucleic acid sequences (SEQ ID NOS: 1, 3, 33, 35, 37, 39, 41, 5, 7, 9, 11 and 13) are used to design oligonucleotide primers for extending a partial nucleotide sequence to full length or for obtaining 5' sequences from genomic libraries. One primer is synthesized to initiate extension in the antisense direction (XLR) and the other is synthesized to extend sequence in the sense direction (XLF). Primers allow the extension of the known cell cycle interacting protein encoding sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers are designed from the cDNA using OLIGO® 4.06 Primer Analysis Software (National Biosciences), or another appropriate program, to be preferably 22-30 nucleotides in length, to have a GC content of preferably 50% or more, and to anneal to the target sequence at temperatures preferably about 68°-72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided. The original, selected cDNA libraries, prepared from mRNA isolated from actively dividing cells or a plant genomic library are used to extend the sequence; the latter is most useful to obtain 5' upstream regions. If more extension is necessary or desired, additional sets of primers are designed to further extend the known region. By following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix, high fidelity amplification is obtained. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR is performed suing the Peltier Thermal Cycle (PTC200; MJ Research, Watertown MA) and the following parameters:

Step 1	94°C for 1 min (initial denaturation)
Step 2	65°C for 1 min
Step 3	68°C for 6 min
Step 4	94° for 15 sec

Step 5	65°C for 1 min
Step 6	68°C for 7 min
Step 7	Repeat steps 4-6 for 15 additional cycles
Step 8	94°C for 15 sec
Step 9	65°C for 1 min
Step 10	68°C for 7:15 min
Step 11	Repeat step 8-10 for 12 cycles
Step 12	72°C for 8 min
Step 13	4°C (and holding)

A 5-10 μl aliquot of the reaction mixture is analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were selected and cut out of the gel. Further purification involves using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc). After recovery of the DNA, Klenow enzyme was used to trim single-stranded, nucleotide overhangs creating blunt ends which facilitate religation and cloning. After ethanol precipitation, the products are redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1 μl T4 polynucleotide kinase are added, and the mixture is incubated at room temperature for 2-3 hours or overnight at 16°C. Competent E. coli cells (in 40 μl of appropriate media) are transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook, supra). After incubation for one hour at 37°C, the whole transformation mixture is plated on Luria Bertani (LB)-agar (Sambrook, supra) containing 2xCarb. The following day, several colonies are randomly picked from each plate and cultured in 150 μl of liquid LB/2xCarb medium placed in an individual well of an appropriate, commerically-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture is transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample is transferred into a PCR array. For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of 4Tth DNA polymerase, a vector primer and both of the gene specific primers used for the extension reaction are added to each well. Amplification is performed using the following conditions:

Step 1	94°C for 60 sec
Step 2	94°C for 20 sec
Step 3	55°C for 30 sec
Step 4	72°C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72°C for 180 sec
Step 7	4°C (and holding)

Aliquots of the PCR reactions are run on agarose gels together with molecular weight markers. The sizes of the PCR products are compared to the original partial cDNAs, and appropriate clones are selected, ligated into plasmid and sequenced.

Example 9: VbDBP(SEQ ID NO: 11)

When a BLAST data base was used it was found that the VbDBP clone is very similar to the putative DNA binding protein (*Arabidopsis thaliana*) and also contains a lot of homologies with PCF2 (*Oryza sativa*). VbDBP interacts with CDC2b but not with CDC2a. The publicly available databases were screened with the cDNA VBDPBP (N-term). With the help of BLASTX gene21 from AC003680 (score 1.0e-27) was found as best homologue. This is a genomic sequence from A.thaliana (entered in the databank:20-MAR-1998), chromosome II. The prediction made here gives 1 big exon, but the new predictions made in accordance with the present invention gave two exons (the big one, followed by a small one). The cDNA VBDPBP shows not so high homology (gene 21 might only be from the same family as VBDPBP) with the big exon, so completion of the cDNA will confirm one or the other annotation and might give a new sequence. Other homologues are D87261) PCF2 [Oryza sativa] (score 9.2e-27) and D87260) PCF1 [Oryza sativa] (score 8.5e-24) both with publication: Kosugi, S. and Ohashi,Y. PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene. Plant Cell 9 (9), 1607-1619 (1997). With the help of BLASTN/nr an other genomic sequence from chromosome V, AB010072 (2e-12) (08-JAN-1998) sequenced by the KAOS-people (P1 clone: MEE6) was found. The region with homology is located between (18754..18848) has no annotations at all. The publicly available databases was screened with the cDNA VBDPBP (C-term (SEQ ID NO: 15)) but nothing was found with BLASTX.

PCF1 and PCF2 are proteins isolated in rice that specifically bind to sites IIa and IIb in the promotor region of the rice PCNA gene (Kosugi et al., 1997). The rice proliferating cell nuclear antigen (PCNA) protein is an auxiliary protein of DNA polymerase (that participates in a variety of processes, such as DNA replication, DNA repair synthesis, and cell cycle control through reactions with the CDK-cyclin-CKI complex. The PCNA gene is induced at the G1-to-S phase boundary and is well conserved in eukaryotes.
The expression of the rice PCNA gene is restricted exclusively to meristematic regions and is controlled at the transcriptional phase. PCNA protein is also present in

proliferating cells but absent from nondividing cells and terminally differentiated plant tissues.

Loss-of-function analysis of the rice PCNA promoter using transgenic plants has demonstrated that two elements (sites IIa and IIb) in the proximal region are essential for the proliferating cell-specific transcriptional activity. On the other hand, two repeated site IIa sequences located upstream of the cauliflower mosaic virus 35S minimal promoter confer transcriptional activation in tobacco protoplast. This suggests that sites IIa and IIb most probably function as positive cis-acting elements in proliferating cells.

The proteins PCF1 and PCF2 specifically bind to sites IIa and IIb in the promoter region of the rice PCNA gene and may act as transcription factors to control DNA synthesis-related genes in plants. In particular, PCF2, with a high level of DNA binding activity in meristematic tissues, may act as transcriptional activator for these genes. These proteins have a deduced basic helix-loop-helix (bHLH) motif that is responsible for DNA binding and dimerization. PCF1 and PCF2 are novel types of bHLH proteins that are distinct from other known bHLH transcriptional factors.

Kosugi, S., and Ohasi Y. (1997) PCF1 and PCF2 specifically bind to cis elements in the Rice proliferating cell nuclear antigen gene. The Plant Cell, 9, 1607-1619.

Example 10: Vb33 (SEQ ID NO: 5)

The Vb33 clone encodes a protein interacting with CDC2b but not with CDC2a. The publicly available databases were screened with the cDNA VB33. With the BLASTX as best homologue a predicted gene on the Z49937 sequence having a similarity with an ankyrin motif (score 0.62) was found. This sequence comes from C.elegans cosmid and the gene F14F3.2 was predicted based on a C.elegans EST (yk192g4.5).

Example 11: LDV115 (SEQ ID NO: 1)

The *LDV115* gene encodes a protein interacting with CDC2a but not with CDC2b and showing limited similarity to the *Saccharomyces cerevisiae* WEB1 protein. The publicly available databases were screended with the cDNA LDV115. With the BLASTX it was found as best homologue the WEB1 protein from S.pombe (AB004537)(score 6.7e-17). This protein as well as the other hits were mainly due to proline-richness of the LDV115 translation. The homology is low but spread over about 50% of the S.pombe protein, which might indicate that LDV115 is at least a member of the family. The *WEB1* gene was isolated as a yeast homologue of the adenoviral E1A gene (Zieler et al., 1995, MCB 15, p3227-3237). The protein products of the E1A gene are implicated in a variety of transcriptional and cell cycle events, involving interactions with several proteins present in the human cells, including parts of the transcriptional machinery and negative regulators of cell division such as the Rb gene product and p107. WEB1 is identical to SEC31, a protein involved in budding of transport vesicles from the endoplasmic reticulum (Pryer et al. (1993), J. Cell. Biol. 120, p865-875). The protein similarity between WEB1 and LDV115 is almost completely due to the presence of a proline-rich region found in both proteins. Proline-rich regions are not restricted to the WEB1 protein, but can also be found in many structural proteins such as hydroxyproline-rich glycoproteins and extensins. Therefore, LDV115 might not be a true homologue of WEB1.